ABSTRACT

Title of Dissertation:	ESTIMATES OF REGIONAL CARBON DIOXIDE FLUXES USING A DENSE NETWORK OF LOW-COST IN SITU OBSERVATIONS
	Cory Ray Martin, Doctor of Philosophy, 2018
Dissertation directed by:	Professor Ning Zeng, Department of Atmospheric and Oceanic Science

Current inverse modeling-based estimates of carbon dioxide (CO_2) fluxes in urban areas typically use a network of 10-20 observation sites featuring high-accuracy gas analyzers that can cost over $100,000 each. Recently, commercially available, low-cost sensors to measure both traditional meteorological quantities and trace gases such as CO_2 have become a focus of atmospheric science research. These flux estimations are an ill-posed problem in the sense that, depending on resolution, the mathematical model may be optimizing fluxes for hundreds or even thousands of grid points, with only relatively few observations to use as constraint. Theoretically, by introducing many more observations into the system, the result will better represent the true state of the surface fluxes.

This work comprises of three related studies that evaluate the viability of using a low-cost CO_2 sensor combined with a mesoscale meteorology model with online tracers, and an advanced ensemble data assimilation technique, to estimate

surface fluxes of CO_2 in an urban region. First, the SenseAir K30 sensor is evaluated compared to a reference gas analyzer to determine the accuracy and precision of the observations from this sensor. Next, a simulation of atmospheric CO_2 is evaluated against observations to understand the error in simulated mole fractions from variations in existing emissions inventories. Finally, a series of observing system simulation experiments (OSSEs) are conducted to understand the sensitivity of estimated CO_2 fluxes to the ensemble data assimilation system configuration.

From this work, it is found that the K30 sensor can be useful for urban ambient monitoring of CO_2 after corrections for environmental factors such as temperature and pressure. Additionally, the modeled CO_2 results show that the error in simulated mole fractions is likely larger from meteorological error than it is from uncertainty in emissions. Finally, the OSSEs find that this ensemble data assimilation system using a dense network of lower-accuracy observations can achieve comparable CO_2 flux estimation results to that of using a sparse network of high-accuracy observations. However, the configuration of the system, particularly the inflation technique used, can significantly affect the quality of the analyzed fluxes.

ESTIMATES OF REGIONAL CARBON DIOXIDE FLUXES USING A DENSE NETWORK OF LOW-COST IN SITU OBSERVATIONS

by

Cory Ray Martin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Ning Zeng, Chair
Professor Russell R. Dickerson
Professor Eugenia Kalnay
Dr. Anna Karion
Professor Jelena Srebric, Dean's Representative
Professor Da-Lin Zhang

Acknowledgements

There are literally dozens of people who have helped me in various capacities over the course of my life to get me where I am today and thanking each one would be a dissertation in itself. However, I would still like to individually thank those who have been essential to the completion of my graduate degrees at the University of Maryland.

First of all, I would like to thank my advisor, Professor Ning Zeng, for all of his help, guidance, and feedback every step of the way during my time here at UMD. Always pushing me to be better, I know that I owe a large part of my success to his support (and I'll always carry with me the importance of properly commented code).

I want to thank Professors Russ Dickerson, Eugenia Kalnay, Jelena Srebric, and Da-Lin Zhang for serving on my dissertation committee and providing valuable feedback and suggestions.

I want to thank all of the scientists that I worked closely with at NIST over the past 3+ years including, but not limited to: Dr. David Allen, Dr. Subhomoy Ghosh, Dr. Sharon Gourdji, Dr. Israel Lopez-Coto, Dr. Kim Mueller, Dr. Kuldeep Prasad, and Dr. Tamae Wong. A special thanks goes to Dr. Anna Karion for not only serving on my committee but also for playing an advisory role in my research as well. Finally, my time at NIST wouldn't have happened if not for Dr. James Whetstone, whom I have learned a great deal from over the past few years on a variety of subjects, most notably how not to make an airtight container.

I want to thank everyone in the Department of Atmospheric and Oceanic Science, especially the front office staff, as they are the glue that holds the department together. A special thanks goes out to Tammy Hendershot, as not only does she always have the back of every single student in our department, but she and I have spent a lot of time over the past 5+ years trying to make AOSC as great as it can be with recruiting and swag, among other things.

There are two other members of the AOSC faculty that I would like to thank individually. First, I want to thank Dr. Tim Canty, for not only being the driving force behind much of the department, but for also giving me the opportunity to teach my first year as a TA for his class, and for all of his help and advice on a variety of subjects along the way. I also want to thank Dr. Daryl T. Kleist, for being a mentor and a friend, and because of his synoptic meteorology expertise, I will always be able to find where the polar jet is.

I wouldn't be where I am today without the support of all of my friends. I want to particularly thank Neal T. Graham and Gina Mazzuca, fellow TAs and friends from day 1 at UMD for commiserating with me, as well as sharing tons of memories. I also want to thank Chris Birchfield and Andy Baumann for providing much needed stress relief in the form of online Xbox games and ridiculous group chats.

Finally, I want to thank my mom, dad, and brother, as well as the rest of my family for all of their support over my entire life.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

2D	Two-dimensional
3D	Three-dimensional
4D	Four-Dimensional
AGL	Above Ground Level
ARL	Observing site in Arlington, VA
BME280	Bosch Sensortech BME280 sensor
CDIAC	Carbon Dioxide Information and Analysis Center
CEAS	Cavity Enhanced Absorption Spectrometry
CEMS	Continuous Emissions Monitoring Systems
CFSR	Climate Forecast System Reanalysis
CRDS	Cavity Ring-Down Spectroscopy
CH_4	Methane
CO_2	Carbon dioxide
DA	Data Assimilation
EDGAR	Emissions Dataset for Global Atmospheric Research
EIA	United States Energy Information Administration
EN	Earth Networks
ESRL	Earth System Research Laboratory
FFDAS	Fossil Fuel Data Assimilation System
FLAGG-MD	Fluxes of Atmospheric Greenhouse Gases in Maryland
GHGs	Greenhouse Gases
HAL	Observing site in Halethorpe, Maryland
H_2O	Water
HVAC	Heating, Ventilation, and Air Conditioning
I^2C	Inter-Integrated Circuit
Infl_Constant	Additive inflation method constant in space
Infl_PtSrc	Additive inflation method constant in space including the locations of higher spread at point sources
Infl_Truth	Additive inflation method using a scaled version of the assumed true emissions
IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
K30	SenseAir K30 CO_2 sensor
KBWI	Baltimore-Washington Thurgood Marshall International Airport
KDCA	Washington Reagan National Airport
KIAD	Washington Dulles International Airport
LETKF	Local Ensemble Transform Kalman Filter
LGR	Los Gatos Research
MAE	Mean Absolute Error
NARR	North American Regional Reanalysis
NCEP	National Centers for Environmental Prediction
NDC	Observing site in northwest Washington, DC
NDIR	Non-dispersive infrared
NEC-B/W	Northeast Corridor – Baltimore/Washington

NIST	National Institute of Standards and Technology
NRT	Near Real-Time
N_2O	Nitrous oxide
NO_x	Nitrous oxides
NOAA	National Oceanic and Atmospheric Administration
O_3	Ozone
ODIAC	Open-source Data Inventory for Anthropogenic CO_2
ODIACFIX	Open-source Data Inventory for Anthropogenic CO_2 without time varying fluxes
OSSE	Observing system simulation experiment
PBL	Planetary boundary layer
PPB	Parts per billion; nanomoles mol^{-1}
PPM	Parts per million; micromoles mol^{-1}
R^2	Coefficient of determination
RCP	Representative Concentration Pathway
RH	Relative Humidity
RMSE	Root Mean Square Error
RPi	Raspberry Pi
SNP	Observing site at Shenandoah National Park
SO_2	Sulfur dioxide
Std. Dev.	Standard deviation
TIMES	Temporal Improvements for Modeling Emissions by Scaling
UART	Universal Asynchronous Receiver/Transmitter
UTC	Universal Coordinated Time
UVA	University of Virginia
VEGAS	VEgetation-Global-Atmosphere-Soil Model
WRF	Weather Research and Forecasting model
WRF-Chem	WRF model coupled with chemistry
YSU	Yonsei University

Chapter 1 Introduction

1.1 Background

Carbon dioxide (CO_2), like methane (CH_4), nitrous oxide (N_2O), ozone (O_3), and water vapor (H_2O), is one of the major greenhouse gases (GHGs) in Earth's atmosphere. While each one of these trace gases is only a small fraction of the total dry atmosphere, especially relative to molecular nitrogen and molecular oxygen, which combined make up nearly 99%, they each have a fundamental importance to our planet's climate. Because of its molecular structure and vibrational frequencies, CO_2 is able to absorb and reemit infrared radiation at many of the wavelengths emitted by the Earth's surface, warming the lower atmosphere. Without it and the other GHGs, our planet would be over 30ºC on average cooler than it is today, making life on Earth less habitable.

After water vapor, the mole fraction of which is driven by thermodynamics and can vary significantly in the troposphere between as little as 0.01% and as high as 4%, CO_2 is the second most abundant GHG on Earth. Since measurements started at the Mauna Loa Observatory on the island of Hawai'i in the 1950s (Keeling et al., 2005), CO_2 has steadily risen from the preindustrial mole fraction of approximately 280 μmol mol^{-1} of dry air (parts per million, or ppm), to approximately 315 ppm when the continuous record began, and now to today's level exceeding 400 ppm as shown in Fig. 1.1. This figure has come to be known as the Keeling curve, which shows the monthly and seasonally averaged CO_2 mole fractions at the Mauna Loa Observatory as a proxy for the globally averaged free troposphere CO_2 mole fraction.

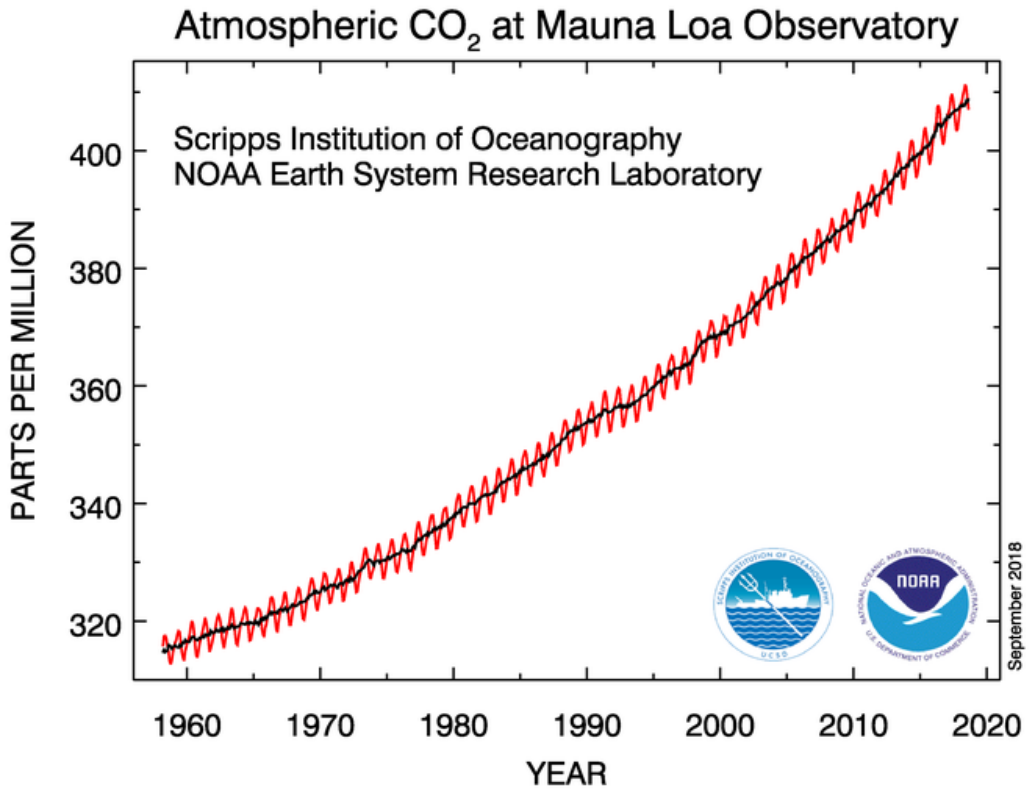

Figure 1.1. Monthly mean (red) in situ atmospheric CO_2 mole fractions observed at the Mauna Loa Observatory from 1958-2018. The annual mean (black) is also plotted to account for the variations associated with the seasonal cycle of CO_2. Figure from NOAA/ESRL; https://www.esrl.noaa.gov/gmd/ccgg/trends/full.html, accessed on September 20, 2018.

Figure 1.2 shows a simplified cartoon of the globally averaged sources, sinks, and pools of carbon on Earth. Annual global carbon emissions from anthropogenic activities has reached over 9 gigatonnes of carbon per year (NOAA ESRL, 2010). While this number is small relative to biological processes or the total amount of carbon stored in the Earth's crust, the global carbon cycle is no longer in balance. Some of this additional carbon is thought to be removed from the atmosphere by photosynthesis and uptake by the ocean, but even with these additional sinks the amount of carbon dioxide in Earth's atmosphere is increasing by four gigatonnes per

year. This accounts for the approximately 2 ppm per year increase in globally averaged CO_2 as reflected in the Keeling curve shown in Fig. 1.1.

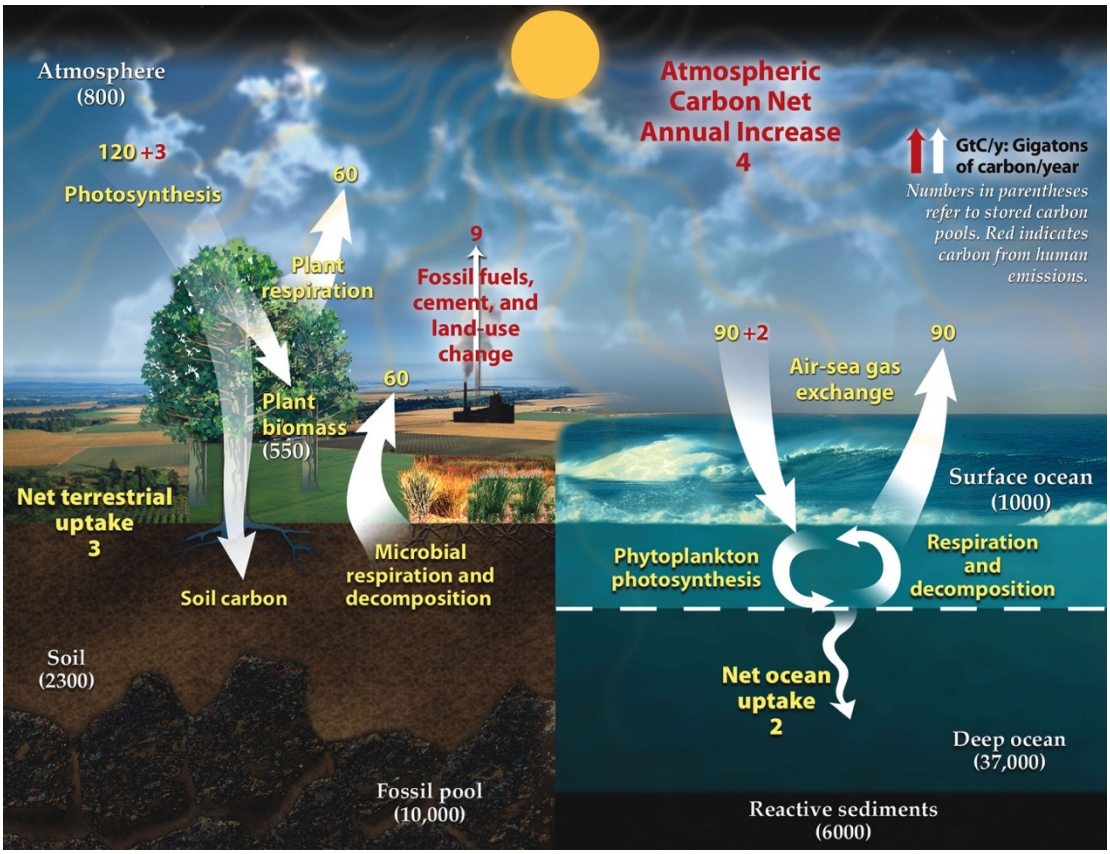

Figure 1.2. Simplified representation of the global carbon cycle. Numbers in yellow and red represent fluxes with units of gigatonnes/year; numbers in parentheses represent stored pools with units of gigatonnes. Figure courtesy of U.S. Department of Energy Office of Science.

While there are many anthropogenic sources of this increase, including biomass burning and deforestation, this growth can largely be attributed to the burning of fossil fuels, which began to become widespread across the globe over the past two to three centuries, and has been the main way our global energy demands have been met ever since. The burning of any organic material, in this case CH_4, can be shown by the following example of a combustion reaction:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O + \text{Heat} \quad (1.1)$$

As is the case for all fossil fuels, which are basically hydrocarbons, when they burn they combine with the oxygen in the air, and in complete combustion produce CO_2 and water vapor, and are exothermic, meaning the give off heat used to drive a turbine or engine to produce electricity, for example. In reality, incomplete combustion often occurs, meaning there is not enough oxygen to react fully with the hydrocarbon, which leads to the production of carbon monoxide (CO). Additionally, combustion of impure fossil fuels also may lead to production of air pollution relevant oxides including nitrogen oxides (NO_x) and sulfur dioxide (SO_2).

To better estimate the future climate as a result of the increased anthropogenic emissions of GHGs, a number of emission reduction goals, also called representative concentration pathways (RCPs; Clarke et al., 2014), have been adopted by the Intergovernmental Panel on Climate Change (IPCC) as part of its fifth assessment report. Each of these RCPs were designed to encompass a range of possible scenarios for the future of anthropogenic CO_2 emissions over the next century: one where emissions peak in the next few years, two others with peaks occuring later in the century, and one where emissions continue to rise through 2100. These four scenarios: RCP2.6, RCP4.5, RCP6.0, and RCP8.5 (named for the additive radiative forcing in W/m^2 associated with the increase in greenhouse gas concentrations) assume a global mean mole fraction of atmospheric CO_2 of 421 ppm, 538 ppm, 670 ppm, and 936 ppm, respectively, by 2100 (IPCC, 2013). For stakeholders and policy makers to assess the success or failure of emissions mitigation strategies, it is

essential to have an accurate assessment of the total emissions from a political jurisdiction or area both before and after a policy is implemented.

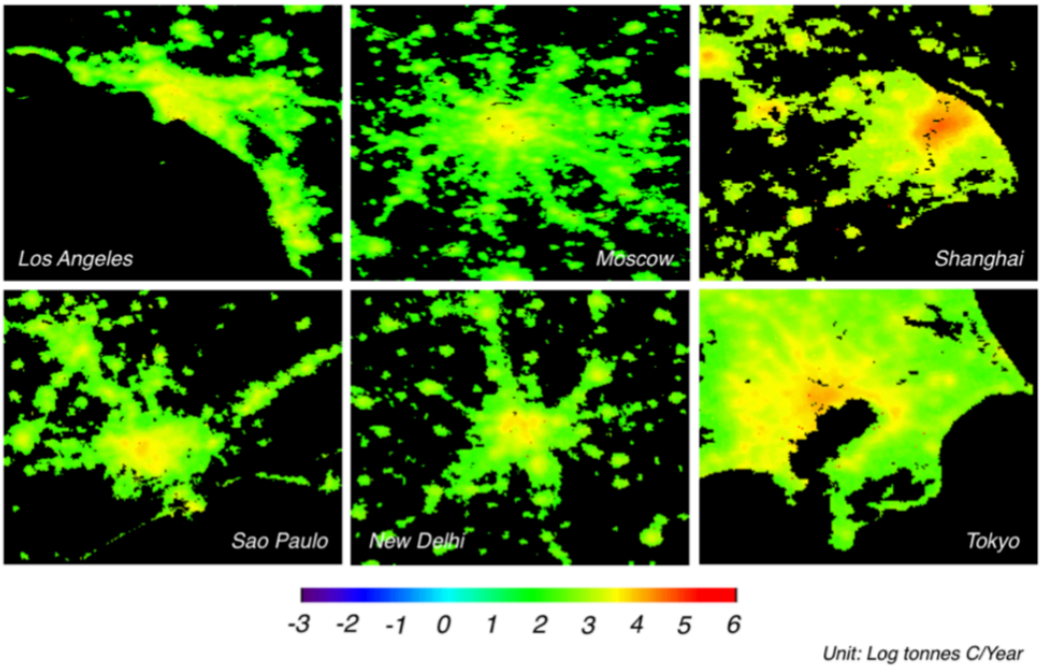

Figure 1.3. An example of a gridded anthropogenic emissions inventory (in this case the Open-source Data Inventory for Anthropogenic CO_2 or ODIAC inventory) showing the magnitude and spatial variability of emissions in six urban areas. Figure from Oda and Maksyutov, 2011.

There are, generally speaking, two main methods for estimating the total emissions from an area of interest, the "bottom-up" approach, and the "top-down" approach (Leip et al., 2018). The bottom-up approach is the more traditional method, where individual source emissions and sinks are estimated based on things such as energy consumption, traffic counts, and actual emissions observations from point sources, and then aggregated either into totals by location and sector or made into emissions inventory products with spatial (and possibly temporal) variability (eg. Gurney et al., 2009; Rayner et al., 2010; Oda and Maksyutov, 2011; Asefi-Najafabady et al., 2014; Oda et al. 2018). An example of a gridded emissions

inventory is shown in Fig. 1.3. Additionally, some studies implement a third method, the mass balance approach, where flux is calculated by the difference of mass upwind and downwind of a theoretical box, generally from aircract observations, but this method can only provide the total flux from the area of interest, and has no discrete spatial information.

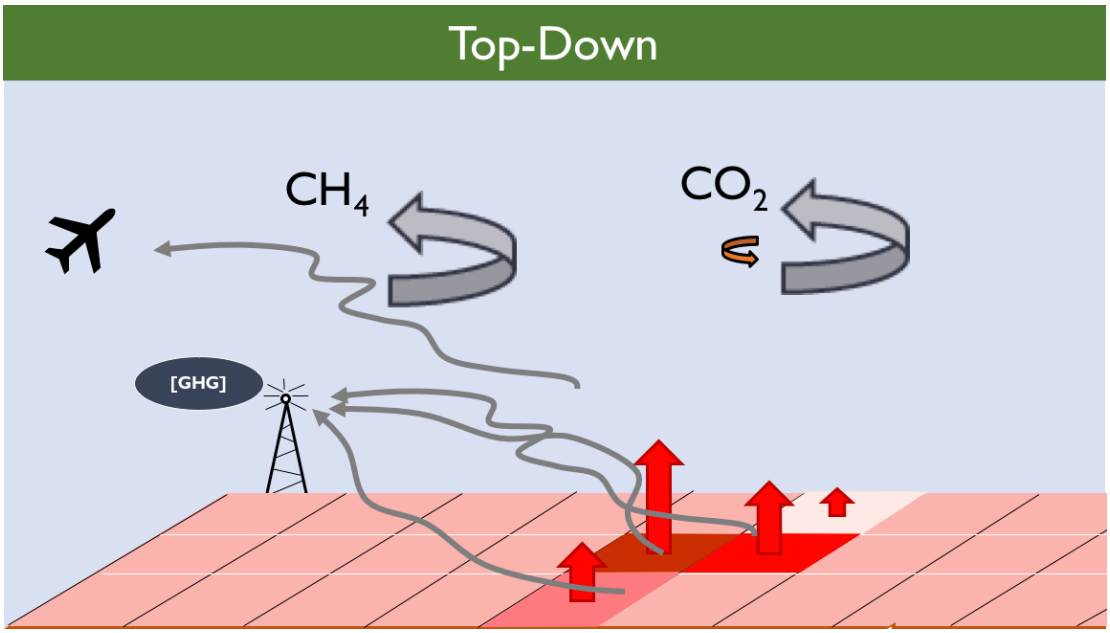

Figure 1.4. An illustration of the top-down method for estimating surface emissions of GHGs including CH_4 and CO_2. Observations of atmospheric mole fractions from a point (a tower or from an airplane) include a component of atmospheric transport as well as the surface fluxes. Figure courtesy of Kim Mueller, NIST.

When estimating emissions with the top-down method an atmospheric transport model is used to generate three-dimensional (3D) wind fields for a time period of interest. In addition to this meteorological model, a Lagrangian particle dispersion model is typically used which is driven by the generated wind fields to determine the area of influence surface fluxes have on an observation for a particular location and time. Then coupled with some prior information on the expected fluxes (often from an emissions inventory created from the aforementioned bottom-up

method), a synthetic observation is created. Once this synthetic mole fraction is generated, it is compared to the observed value in order to optimize the fluxes for the area of influence using one of many techniques to minimize the error between the two mole fractions. This has been done extensively, using numerous optimization techniques, for surface carbon flux estimation both at the global scale (e.g. Bousquet et al., 1999; Michalak et al., 2004; Gourdji et al., 2008; Mueller et al., 2008) as well as at the regional or urban scale (eg. Gerbig et al., 2003; Peylin et al., 2005; McKain et al., 2012; Lauvaux et al., 2016). A diagram illustrating this technique is shown in Fig. 1.4.

Methods of numerical weather prediction, atmospheric data assimilation techniques such as variational methods and Kalman filters (Peters et al., 2005; Baker et al., 2006; Kang et al., 2011, 2012; Chatterjee and Michalak, 2013; Liu et al., 2016) have been used to estimate surface fluxes. For these, synthetic observations of CO_2 are generally created from interpolation of output of online tracer transport models, rather than the Lagrangian particle dispersion method described earlier, but not necessarily so. The former mimics the data assimilation methodology used for traditional meteorological variables such as temperature and atmospheric pressure.

Half of the world's population now lives in cities, and the United Nations expects this trend to continue to two-thirds by 2050 (United Nations, 2014). The majority of anthropogenic emissions are also from urban areas, where electrical and industrial energy generation and transportation are the largest emissions sources located in these geographically small but densely populated areas. In recent years, several urban GHG measurement campaigns have been implemented to improve

measurement and quantification of greenhouse gas concentrations in, as well as the fossil fuel emission fluxes from metropolitan areas. These cities include but are not limited to Salt Lake City (McKain et al., 2012), Boston (Briber et al., 2013), Indianapolis (Turnbull et al., 2015; Lauvaux et al., 2016; Miles et al., 2017), Paris (Breón et al., 2015), and Los Angeles (Kort et al., 2013; Feng et al., 2016). Generally, these campaigns feature a relatively dense network of towers equipped with inlets at a height above the ground leading to state-of-the-art cavity ring-down spectroscopy (CRDS) GHG analyzers with sampling systems and mole fraction standards for periodic calibration. The observations from these towers placed upwind, downwind, and inside the metropolitan areas are then used to validate a prior bottom-up emissions inventory as well as generate an optimized top-down emissions estimate. The most recent urban GHG measurement campaign, led by the National Institute of Standards and Technology (NIST) is for the Northeast Corridor of the United States, with the first part of the network beginning in the region around the Baltimore, Maryland and Washington, District of Columbia metropolitan areas (Lopez-Coto et al., 2017; Mueller et al., 2018).

CO_2 observations, both from flask samples and state-of-the-art continuous measurement instruments, have a typical compatibility goal of ~0.1 ppm, recommended for observations at background global network sites (World Meteorological Organization, 2013). Flask-based measurements require observers to collect samples for lab analysis, at significant cost. Continuous in-situ CO_2 analyzers located at towers or on top of buildings do not suffer from these regular costs, but these high-precision analyzers can cost upwards of $100,000 per site, plus any

additional costs for calibration gases and installation of equipment and inlet lines. High-accuracy CO_2 observations are thus relatively sparse compared to other climatological variables such as temperature and precipitation. Even for the aforementioned urban campaigns, where the density of CO_2 observations is much higher than the rest of the world, these urban areas, which each can be over 10,000 km^2, still only have approximately 10-16 measurement sites.

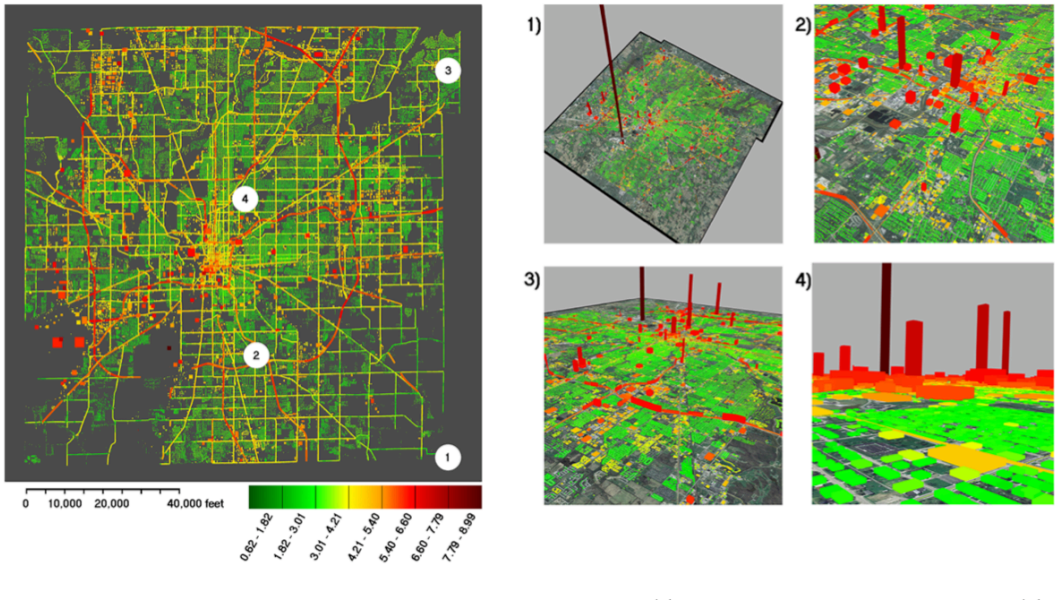

Figure 1.5. Total fossil fuel CO_2 emissions for Indianapolis, IN at the building/street level from the Hestia project (Gurney et al., 2012). Color bar is in units of log_{10} kg C/year.

This high accuracy is needed on regional or global scales where the variations in CO_2 can be relatively small, but in urban areas the diurnal and synoptic variability in near-surface CO_2 may be as high as 25% of the total observed mole fraction. This is reflected in the heterogeneity of urban emissions, where efforts are ongoing to estimate fluxes at the sub-kilometer scale, or even for individual city blocks and buildings, an example of one such very high-resolution emissions inventory for Indianapolis is shown in Fig. 1.5 (Gurney et al., 2012). For these urban applications

where enhancements of CO_2 relative to the global background average are sufficiently large, uncertainty in the observations can still be as large as 1% and provide useful information. Observing system simulation experiments (OSSEs) have found that, depending on the methodology used, a higher spatial density of observations in these urban regions has been shown to better constrain the inversion estimates, even if the absolute uncertainty of the observations is higher (Turner et al., 2016; Wu et al., 2016; Lopez-Coto et al., 2017; Wu et al., 2018), but a trade-off between total network cost and inversion constraint must be balanced.

In recent years, a wave of small, low-cost sensors, some of which measure trace gases or particulate matter, in addition to traditional meteorological variables, using various technologies have become commercially available. Evaluation and implementation of some of these new low-cost sensors demonstrate their promise for ambient air monitoring (Eugster and Kling, 2012; Holstius et al., 2014; Piedrahita et al., 2014; Young et al, 2014; Wang et al., 2015; Shusterman et al., 2016; Lewis et al., 2018). Many of these instruments are based on electrochemical reactions to measure the concentrations of trace gases. With the advent of widely available and low-cost mid-infrared light sources and detectors, a small group of non-dispersive infrared (NDIR) CO_2 sensors have also become commercially available. They are designed for use in a number of applications including ventilation control, agricultural and industrial applications, and inclusion in stand-alone commercial products. Additionally, with the high volume of possible applications, these small NDIR CO_2 sensors are affordably priced on the order of $100 to $200 per sensor. Previous studies have compared some of these NDIR CO_2 devices and concluded that after

application of some type of calibration procedure, some of these devices can provide reasonably accurate measurements (±3-5ppm) of ambient CO_2 concentrations (Hurst et al., 2011; Yasuda et al., 2012; Shusterman et al., 2016; Kunz et al., 2018; Lewis et al., 2018).

1.2 Thesis Objectives and Outline

In this study, the viability of the use of a dense network of lower-accuracy observations of CO_2 mole fractions for carbon flux estimation from an urban area is explored. Previous work has been done to estimate urban GHG emissions using top-down approaches but only with a relatively sparse network of high-accuracy observations. As the nations and cities of the developing world grow their economies and their associated energy needs increase, the importance of quantification and mitigation of GHG emissions plays an even larger role in the political and environmental stability of our planet as a whole. Specifically, this work aims to answer the following questions:

- Are any of the current commercially available and low-cost observing platforms for CO_2 able to resolve the ambient variability both in time and space with sufficient accuracy and precision for use in urban environments?
- Are high-resolution forward modeled simulations of atmospheric CO_2 mole fractions able to accurately resolve the mean and temporal variability found in observed time series?

- Is the error associated with differences in emissions inventories greater than the error from meteorological transport and dispersion when comparing simulated CO_2 to observations?
- Can an ensemble of forward tracer model simulations be used with in situ observations of atmospheric CO_2 and data assimilation techniques to estimate high resolution surface fluxes of carbon in urban areas?
- How does the estimation of surface fluxes using a dense network of lower accuracy observations compare to using a sparse network of high accuracy observations?
- Is there any added benefit to the surface flux estimation by creating a hybrid network containing both the high accuracy and low accuracy observations over just using the sparse high accuracy observation network?
- How do changes in the configuration of the data assimilation system affect the surface flux estimates?

This dissertation is divided into five chapters. The first (and current) chapter provides an overview of global atmospheric CO_2 and its significance to climate change, as well as a brief description of previous and ongoing efforts to quantify urban GHG emissions. Chapter 2 describes the evaluation and enhancement of the performance of a low-cost CO_2 sensor for use in urban ambient monitoring. A month long co-located experiment with a high-accuracy instrument determined that after correcting for environmental variables in a multivariate linear regression, namely atmospheric pressure, temperature, and water vapor mixing ratio, these sensors could

be used for relatively accurate quantification of urban CO_2 mole fractions. The results of this study led to a paper that was published in *Atmospheric Measurement Techniques* (Martin et al., 2017).

Chapter 3 describes a month-long simulation of atmospheric CO_2 for the region around Baltimore, MD and Washington, DC. Five separate anthropogenic emissions inventories are used as distinct estimated CO_2 enhancements, along with the same background CO_2 concentration applied to each along with a coupled biospheric model for the vegetation related fluxes. The modeled time series are compared to observations from four locations during this period, and while the model and observations agree well on average, there can be significant differences for any given location or time. These differences are determined to be a result of atmospheric transport errors, or other meteorological error and the error caused by the model meteorology is larger than the error caused by variations between the different anthropogenic emissions inventories. This work led to a paper submitted to *Atmospheric Environment* (Martin et al., under review).

Chapter 4 combines the work of Chapters 2 and 3 by using an advanced ensemble data assimilation technique along with a numerical weather prediction model to demonstrate the potential to estimate the fluxes of CO_2 from the Baltimore, MD and Washington, DC metropolitan areas in a series of perfect model OSSEs. In this chapter, the same model from Chapter 3 is used as the transport model along with one emissions inventory to create pseudo-observations of CO_2 mole fractions for four scenarios: 1) a network of 20 high-accuracy tower sites, 2) a network of 200 low-cost sensors with performance similar to what was determined in Chapter 2, 3) a

combination of both types of observations, and 4) an idealized network featuring one high-accuracy observation in each model grid point. Sensitivity of the results to the different observation networks, different ensemble inflation techniques, and observation localization radii is evaluated and presented. The results of these OSSEs are featured in a paper in prep for *Atmospheric Chemistry and Physics* (Martin et al., in prep). Finally, the main conclusions of this dissertation as well as directions for further research are featured in Chapter 5.

Chapter 2 Evaluation and Environmental Correction of Ambient CO_2 Measurements from a Low-Cost NDIR Sensor

Chapter based on manuscript published in Atmospheric Measurement Techniques, July 2017.

Cory R. Martin[1], Ning Zeng[1,2], Anna Karion[3], Russell R. Dickerson[1,2], Xinrong Ren[1,4], Bari N. Turpie[1], Kristy J. Weber[1,5]

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA
[2]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA
[3]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[4]Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD 20740, USA
[5]Now at: Department of Geography, University of Colorado at Boulder, Boulder, CO 80309, USA

2.1 Introduction

Chapter 1 mentioned that a number of small, low-cost sensors, some of which measure trace gases or particulate matter, in addition to traditional meteorological variables have become commercially available, including a small group of NDIR CO_2 sensors. In this chapter, one of these small NDIR CO_2 devices is assessed by determining its accuracy with and without environmental corrections. Section 2.2 describes the CO_2 sensor and its Allan variance, the other instruments included in the system, and the data collection and processing methodology. Section 2.3 describes the calibration and shows the stability of the reference high-precision gas analyzer, and the initial results from the NDIR sensor are shown in Sect. 2.4. In Section 2.5, two methods are described to determine functional relationships and coefficient values to correct the observed values of the instrument for environmental variables and Sect.

2.6 discusses the potential utility of observations from this sensor after correction and temporal averaging.

2.2 Instruments and Methods

To test the validity of using low-cost sensors for scientific applications, a sensor package was implemented consisting of various off-the-shelf components. The K30 sensor module (K30) from SenseAir (Sweden), is the low-cost NDIR CO_2 observing instrument used in this study[1]. The K30 is a microprocessor-controlled device with on-board signal averaging, has a measurement range of 0 to 10,000 ppm, observation frequency of 0.5 Hz, and resolution of 1 ppm. The manufacturer's stated accuracy of the K30 sensor is ±30 ppm ±3 % of reading (SenseAir, 2007) for the 0.5Hz raw output. Additional NDIR sensors were initially evaluated before selecting the K30, including the COZIR ambient sensor and Telaire T6615, which have manufacturer specified accuracies of ±50 ppm ±3 % and ±75 ppm respectively (Gas Sensing Solutions, 2014; General Electric, 2011). The K30 was chosen not only because it has the highest manufacturer-specified accuracy, but also because initial testing showed reliability and consistency when compared to higher-quality observations. In addition to CO_2, temperature, relative humidity, and pressure readings are recorded using a breakout board purchased from Adafruit. This board features a Bosch Sensortec BME280, which according to the manufacturer's datasheet has an average absolute accuracy of ±1 ºC, ±3 %, and ±1 hPa, and an output resolution of 0.1 ºC, 0.008 %, and 0.01 hPa for temperature, relative humidity, and pressure, respectively (Bosch Sensortec, 2015).

To compare the performance of the K30 to better-performing research instrumentation, a greenhouse gas analyzer based on cavity enhanced absorption spectrometry (CEAS) was used as the control. The LGR-24A-FGGA fast greenhouse gas analyzer from Los Gatos Research (LGR, San Jose, CA) provides CO_2, CH_4, as well as water vapor mixing ratios at a frequency of 0.5 Hz and has an un-calibrated uncertainty of < 1 % (Los Gatos Research, 2013). The LGR was connected to a tee connection, to allow either ambient air or a calibration source (during calibrations) to be sampled continuously by the analyzer at a flow rate of 400 standard mL min^{-1}. Calibrations for CH_4 and CO_2 were conducted using several NIST-certified standard mixtures every 23 to 47 hours for a period of one month with molar mixing ratios ranging from 1869.6 parts per billion (ppb) to 2159.4 ppb for CH_4 and from 369.19 ppm to 429.68 ppm for CO_2. See Sect. 2.3 for details and results of this calibration period.

It is important to note that there are differences in how CEAS works compared to NDIR, most notably that the LGR and other CEAS instruments have a controlled cavity where pressure and temperature are kept nearly constant (with a standard deviation of under 0.5 torr and 0.1 ºC for 2-second data), removing potential environmental interference and the need for corrections, whereas the NDIR K30 works in the ambient environment without any mechanism for keeping temperature or pressure constant. Additionally, the LGR implements a water vapor correction on its greenhouse gas concentrations to estimate the dry gas mixing ratio, while the K30 makes no water vapor corrections. A difference between the two analyzers with regard to their sensitivity to the isotopes of CO_2 is expected to be small because the

standards used to calibrate the LGR account for all CO_2 isotopes. To increase the effective path length, both the K30 and LGR use mirrors, but the LGR system uses highly reflective mirrors that allow for an effective path length that is many times longer than that of the K30. Additionally, the CEAS instrument determines the concentration of a gas by how long it takes for the signal to degrade inside the cavity (the e-folding time), whereas an NDIR sensor merely measures the intensity of the signal received relative to the total intensity emitted.

For data collection, a Raspberry Pi (RPi) computer is used (Raspberry Pi Foundation, 2015). The RPi is a credit card sized (approximately 6 x 9 cm) computer running a full Linux distribution, allowing for easy customization and usability, that is priced at around $25. The K30 is connected to the RPi over Universal Asynchronous Receiver/Transmitter (UART) Serial, and the BME280 over Inter-Integrated Circuit (I^2C) serial. An image of the complete sensor package is available in Fig. 2.1. Data is archived on the RPi and uploaded to a centralized data storage and processing server. The LGR collects and archives its own data, but an RPi is used here as well to collect the data from the LGR over a local area network and transfer it to the same centralized server. The added computational power of a Raspberry Pi over traditional data loggers allows for the ability to archive two levels of data: the raw data collected every two seconds, and one-minute averages.

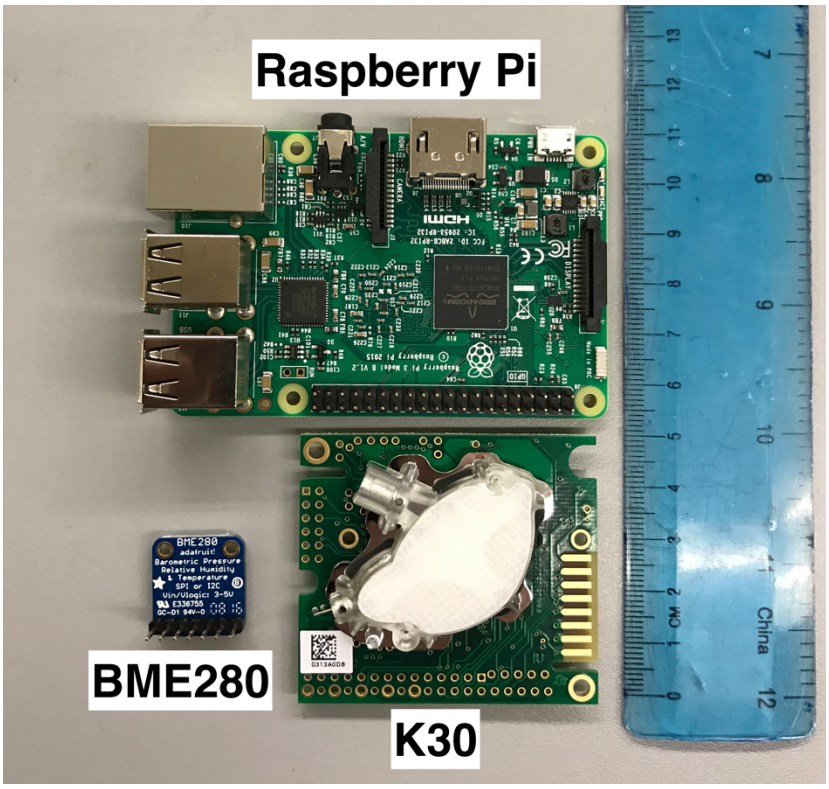

Figure 2.1. Photograph of a Raspberry Pi computer (top), a SenseAir K30 (NDIR) CO_2 sensor (bottom center), a Bosch BME280 temperature and pressure sensor (bottom left), and a ruler for size reference.

Archiving and comparing multiple datasets proved to be challenging, so steps are taken to ensure that each compared value is at the same observed time. All of the RPis use an internet server to synchronize their time, and the LGR uses an internal clock with battery that was set to the same time as the RPis at the beginning of the experiment. Because of various complications including the exact LGR start time and the potential for delays in the RPi's Linux operating system, the data collection times of each K30 sensor package and the LGR are asynchronous. Additionally, power issues can corrupt parts of the plain text data files stored on the RPi's SD card with random characters. Thus, a post-processing procedure has been developed that filters extraneous characters, and then each dataset is synchronized based on recorded time

stamps and averaged over selected time periods. These new datasets can then be directly compared without missing or out of phase data points.

2.2.1 K30 Allan Variance

Allan variance (Allan, 1966) is a measure of the time-averaged stability between consecutive measurements or observations, often applied to clocks and oscillators. In addition, an Allan variance analysis can be used to determine the optimum averaging interval for a dataset to minimize noise without sacrificing signal. Figure 2.2 shows the Allan deviation (the square root of the variance) for one K30's raw two-second data when exposed to a known reference gas. The original two-second data shows the maximum noise, with a standard deviation comparable to the manufacturer's specifications of ±30 ppm, but averaging for even ten seconds drops the variance significantly. According to this analysis, the optimum averaging time, when the Allan variance is at a minimum (Langridge et al., 2008), is approximately three minutes; longer averaging times do not reduce the noise. The other sensors were found to perform similarly. For the subsequent analysis, an averaging time of one minute is used, as the Allan variance is only slightly higher than for three minutes, and one minute observations allow for resolution of atmospheric variability at shorter time scales.

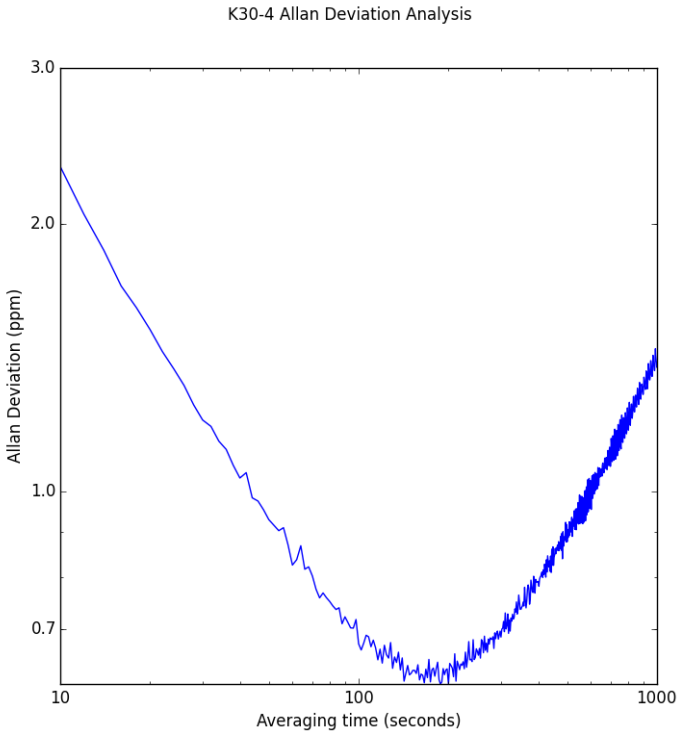

Figure 2.2. Allan variance analysis for an NDIR (K30) CO_2 sensor when introduced to breathing air from a high-pressure cylinder of a constant and known CO_2 concentration. Averaging times between 10 and 1,000 seconds are shown. The black line (slope -0.5) shows where the noise is white or Gaussian. Averaging times greater than about 200 s produce no improvement.

2.2.2 Co-located Experiment

The need to quickly and effectively evaluate a relatively large number of sensors under conditions with relatively stable CO_2 led to the use of a rooftop observation room on the University of Maryland campus in College Park, Maryland. Because this rooftop room had limited access, and it was not part of the building's HVAC system, it served as an ambient evaluation chamber with minimal influence from human respiration. The room was slightly ventilated for the entire evaluation period to allow outside air to slowly diffuse into the room, with a small household box fan also in the room to ensure that the air was well mixed. The room also features

a small, independent heating and cooling unit, but it was only used to keep the room from exceeding a certain temperature, thus the room was not fully temperature controlled. Even with this control, the diurnal fluctuations of temperature in the room were similar to that of the outdoor environment. This ventilation strategy was intentional so that the room then mimicked the ambient CO_2 concentration of the surrounding atmosphere, and approximated the outdoor temperature and humidity, while protecting instruments from direct sunlight, extreme temperatures, and inclement weather. This provided an advantage over controlled tests in a laboratory setting in that rather than just a multi-point calibration, comparing datasets over ambient concentrations and environmental conditions allowed for a realistic evaluation of these instruments in more real-world scenarios.

For a continuous period of approximately four weeks in spring 2016, six K30 sensor packages as described in Sect. 2.2 were deployed alongside the LGR in the rooftop room, all sampling room air. The LGR was also connected to a mass flow controller and standard tank to periodically provide a reference for stability (details in Sect. 2.3). For the reference dataset, the dry CO_2 ($CO_{2\ dry}$) output calculated by the LGR was used. This output includes an applied correction to the mole fraction of CO_2 to give the dry air mole fraction in ppm. The raw CO_2 values were recorded from each K30, temperature and pressure were recorded from each BME280 sensor, and water vapor mole fraction was also recorded by the LGR. All of the observations were recorded every two seconds, and averaged into one minute values. The next two sections describe the stability of the LGR as well as the initial comparison between the K30 and LGR observations.

2.3 Los Gatos Evaluation and Correction

To evaluate the K30 NDIR sensor performance compared to a research-grade analyzer, first the control dataset needs to be calibrated and corrected for drift. To calibrate the LGR, after the experiment concluded the dataset was corrected using a two-point calibration curve derived from using two NIST-traceable gas standards, one with a CO_2 mole fraction of 369.19 ppm, and the other with a mole fraction of 429.68 ppm. A linear fit was then assumed between the two calibration points, with the recorded values as the dependent variable and the NIST-assigned tank values as the independent variable. In addition, three cylinders of breathing air with higher CO_2 mole fractions of 449.73, 486.53, and 516.41ppm (that are NIST-traceable) were also previously used to calibrate the LGR and showed its linearity. Once the coefficients were determined, the entire LGR dataset was then corrected for further analysis.

In addition to the calibration described above, there was a need to quantify any drift in the LGR analyzer. During the experiment period, the LGR was attached to a tee connector, which pulled ambient air from the aforementioned evaluation chamber using its included pump most of the time, but received periodic calibration every 23 to 47 hours for a period of one hour, initially, and later, ten minutes, to conserve the tank, using a reference tank of breathing air connected to a Dasibi Model 5008 calibrator, which was used to schedule the input of calibration gas. This breathing air tank is assumed to have a fixed CO_2 mole fraction, which was estimated by using the LGR to be 463.7 ppm and was used to quantify and subtract the drift of the LGR over the comparison period.

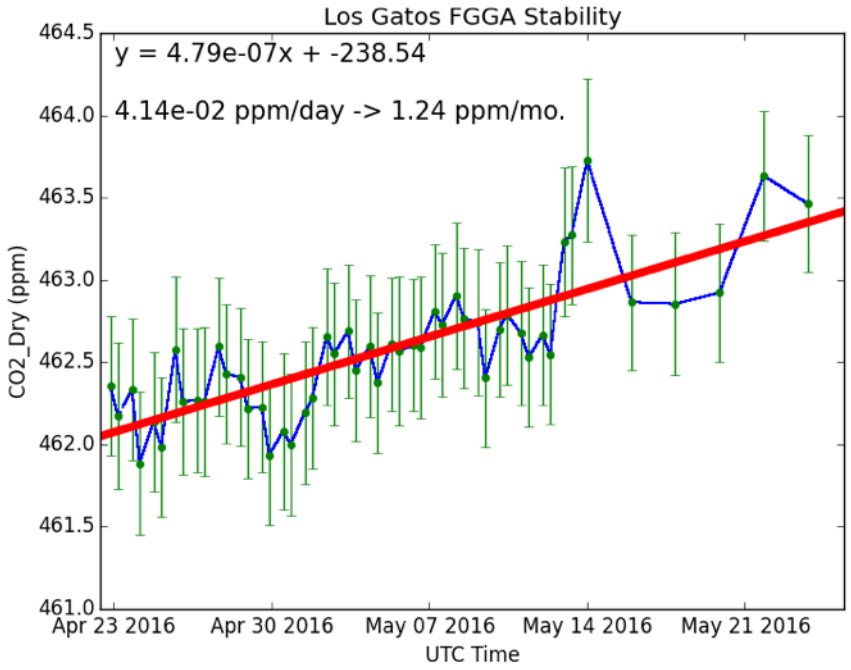

Figure 2.3. Stability of the Los Gatos Fast Greenhouse Gas Analyzer shown over a 30-day period. Excess breathing air with a fixed CO_2 concentration was introduced periodically using a mass flow controller. The mean of each calibration period is plotted in green with the standard deviation as error bars. The blue line is the linear interpolation between each calibration point, and the red line is a linear fit of each calibration point over the entire time series. The red line is subtracted from the dataset to account for the drift of the analyzer over this period.

In Fig. 2.3, the ambient data from the LGR has been filtered out to show only each calibration period performed during the month long experiment. The data during each calibration period was averaged (either a total of 10 minutes or one hour depending on the calibration period) and the averages are plotted on Fig. 2.3. While there is some small variation in the mean mole fraction observed during each calibration from day-to-day, there was an upward trend in the recorded value, by over 1.2 ppm over a 30-day period. This observed drift, while not insignificant, is well within the manufacturer's specifications for this analyzer. However, the observed standard deviation of the two-second points used in each average (the error bars on Fig. 2.3) remained relatively constant throughout the period with a mean standard

deviation of ±0.3 ppm, which is the manufacturer's specified repeatability for 2-second data. This high-frequency noise is not a problem for the analysis with the K30 sensor because both datasets are averaged to one minute values, which removes most, if not all, of this noise. For comparisons between the K30s in the remainder of this paper, the LGR drift is corrected by first computing a linear fit to the calibration points in time (red line, Fig. 2.3) and then subtracting from the LGR dataset the difference of this fit line from the tank's assigned value of 463.7 ppm. After this linear correction, the means of each calibration had an RMSE of 0.2 ppm from the fit line.

2.4 Initial K30 Results

Figure 2.4 shows the original time series of data recorded during the evaluation experiment described in Sect. 2.2.2. The top panel shows raw CO_2 mole fractions reported by six K30 sensors as well as the LGR analyzer, each of which is located in the same rooftop evaluation chamber. The middle panels show the reported atmospheric pressure and temperature values from one BME280 sensor, and the water vapor mole fraction from the LGR. Then, the bottom panel is the difference between the original recorded K30 value and the corrected LGR recorded CO_2 mole fraction with the calibration periods removed.

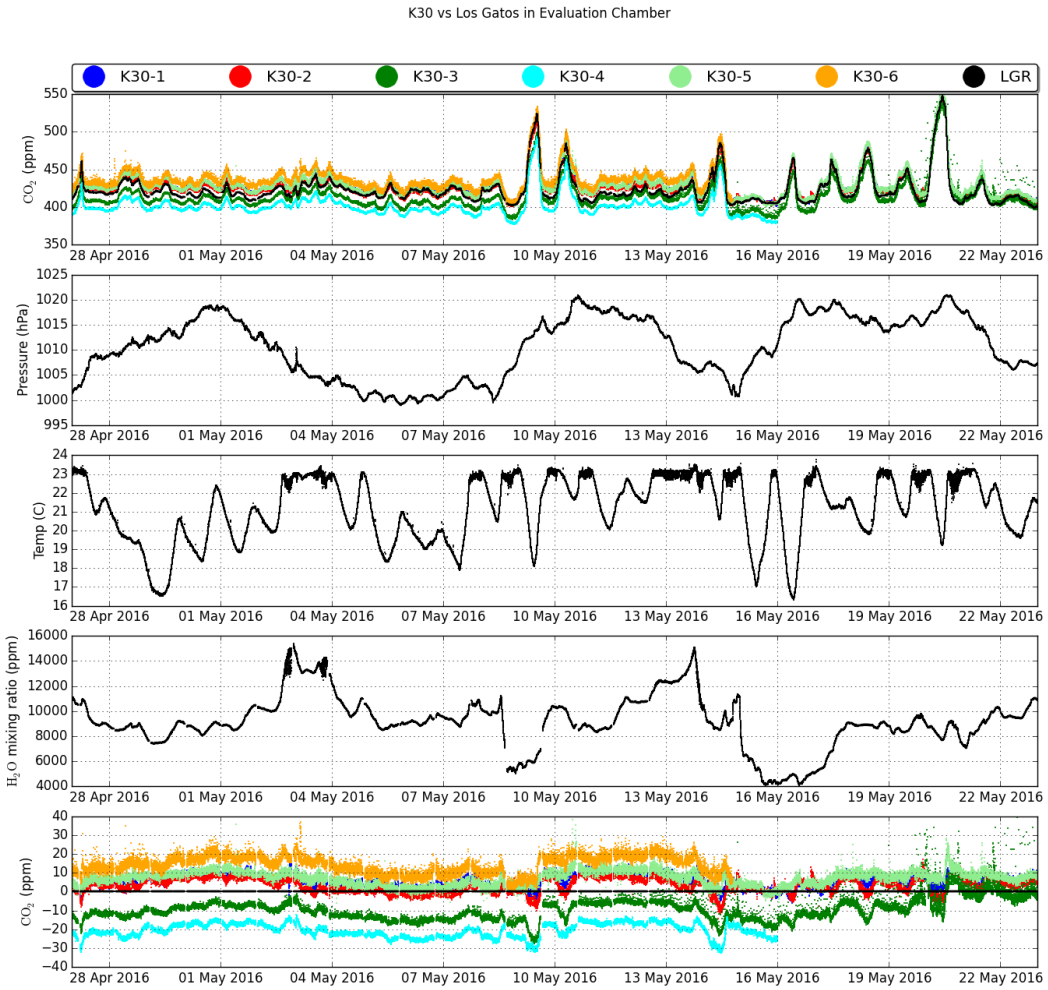

Figure 2.4. Continuous 1-minute time series data during the evaluation experiment. Top panel: CO_2 observed by six K30 sensors as well as the Los Gatos Research Fast Greenhouse Gas Analyzer. Middle panels: observed atmospheric pressure, temperature, and water vapor mixing ratio, respectively. Bottom panel: difference of each K30 from the Los Gatos instrument.

Over this four-week period, the LGR observed an ambient variation of CO_2 with an average value of just over 423 ppm, and a standard deviation of just under 21 ppm. There is distinct synoptic variation in the diurnal cycle observed, with the magnitude varying from as little as 10 ppm over 24 hours to more than 100 ppm. Each of the K30s was successfully able to resolve the ambient variations in CO_2 over this evaluation period, although none of the K30s matched the LGR perfectly in both

absolute concentration and relative change. However, without any correction or calibration, each K30 was well within the manufacturer's stated uncertainty of ±30 ppm ±3 % of the reading for 1-minute values.

From the difference plot (Fig. 2.4, bottom panel), there are some important things to note. First and foremost, each individual K30 sensor has a distinct zero offset. A few of the sensors are approximately the same as the LGR, but many can have an offset that is as much as 5 % (20 ppm) from the LGR. The differences between each K30 and the LGR all have standard deviations between 4 ppm to 6 ppm and root mean square errors (RMSE) between 5 ppm to 21 ppm. This means that after accounting for the offset of each individual K30, the practical accuracy of the K30 CO_2 sensor can be within 1 % of the observed concentration. Secondly, each K30 difference time series appears to feature two wave patterns, one with a period of around one week, and another with a period of approximately one day. Given that the cycles seem fairly consistent and are present in each K30, this suggests that the difference between the recorded values from the LGR and each K30 is not random, but instead that there are external factors that can be assessed for potential compensation in the K30 response.

2.5 Environmental Correction

In Fig. 2.4, the difference between the LGR and each K30 is shown in the bottom panel below time series of environmental data from the evaluation chamber. Just like in the difference plot, each of the environmental variables features two distinct time scales of variability. There is a diurnal cycle of each variable, as well as

synoptic-scale variability attributed to weather systems that occurs on the order of one week. Because the observed CO_2 differences and the environmental variables are correlated on both short and long time scales, statistical regression methods were used to correct the observed concentration of CO_2 from the K30 sensor to a value approximately that of the concentration determined from the calibration-corrected LGR measurements. Generally, a multivariate linear regression is of the form:

$$y = a_1x_1 + a_2x_2 + \cdots a_nx_n + b + \varepsilon_n \quad (2.1)$$

In this case, the measured value *y* is influenced by: the 'true' CO_2 value (taken as the value from the LGR instrument), pressure, and other environmental variables as the dependent variables $x_1, x_2, x_n,$ respectively. A multivariate regression analysis can then be used to find the corresponding coefficients. In addition, in order to better identify the contribution from each individual factor, the data were also analyzed in a successive regression analysis, as described below.

2.5.1 Successive Regression Method

Each individual K30 sensor's original observed CO_2 dataset is first regressed to the LGR dry CO_2 dataset. This regression accounts for the traditional zero and span corrections made during an instrument calibration. The calibration curve of one K30 for just zero and span is shown in Fig. 2.5. But to include biases due to environmental factors, then the residual, epsilon (ε), is calculated as:

$$\varepsilon = y - ax - b \quad (2.2)$$

where in this instance *x*, the independent variable, is the LGR dataset and *y*, the dependent variable, is the K30 dataset.

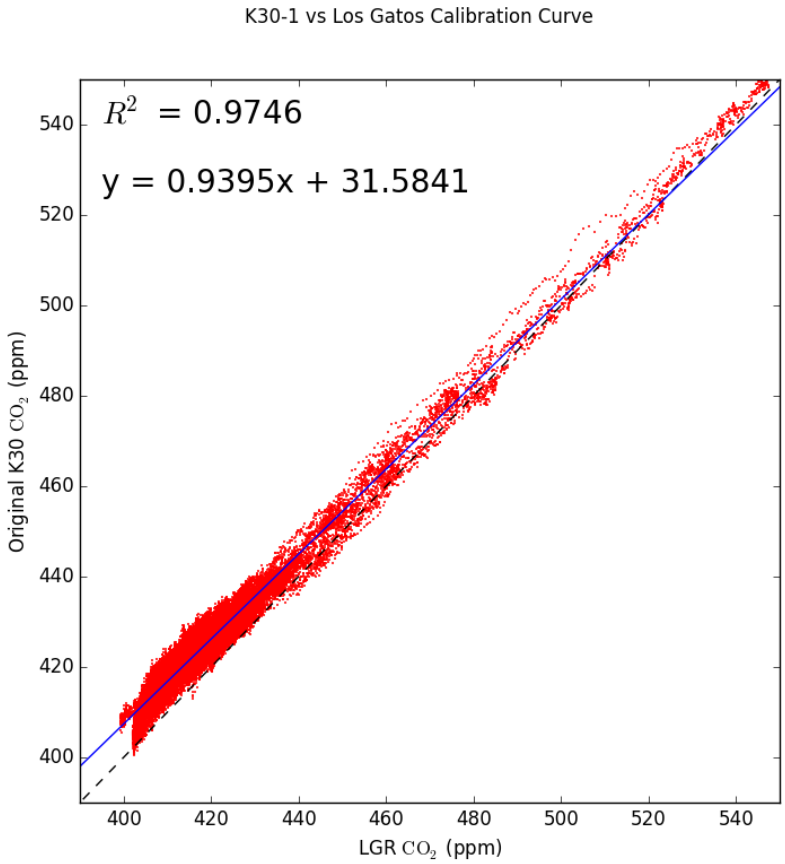

Figure 2.5. Calibration curve of K30-1 vs LGR for 1-minute averages without any environmental correction, only span and zero offset are corrected. Solid line is the best fit; dashes represent the 1:1 line.

This process is repeated for each environmental variable pressure (P), temperature (T), and water vapor (q), where (P,T,q) is the independent variable, x, and the ε from the previous step is the dependent variable, y. This linear regression method leads to eight correction coefficients, of the form a_n and b_n, where n is from 0 to 3 representing each of the independent variables included in the regression. These coefficients can then be used in the following equation along with the environmental variables to correct K30 CO_2 observations for environmental influences.

$$y_{corrected} = \frac{y - b_0 - (a_1 x_1 + b_1) - (a_2 x_2 + b_2) - \cdots - (a_n x_n + b_n)}{a_0} \quad (2.3)$$

Table 2.1. Root mean square error in ppm between the CEAS LGR and each K30 NDIR sensor's one-minute averaged data for: the original dataset before correction, at each step of the successive regression correction (correcting for 1. zero/span, 2. atmospheric pressure, 3. temperature, and 4. water vapor mixing ratio), and after the multivariate regression correction. Each value shown is for a regression calculated using data from the entire evaluation period.

	Original	Zero/Span	Pressure	Temp	q (final)	Multivariate
K30 # 1	6.9	3.3	2.7	2.7	2.1	1.8
K30 # 2	5.4	3.5	2.2	2.2	1.9	1.7
K30 # 3	10.9	6.0	5.0	4.9	4.5	4.3
K30 # 4	20.8	3.7	2.5	2.4	1.9	1.7
K30 # 5	8.3	3.7	2.6	2.6	2.2	2.0
K30 # 6	15.2	4.9	3.6	3.5	2.7	2.2

For one typical K30, the initial standard deviation of the difference between the K30 and LGR, the RMSE of the data was 6.9 ppm. Using the cumulative univariate regression method described above for the entire evaluation period, the RMSE decreased after each step. After the span and offset regression, it dropped significantly to 3.3 ppm. Then after correcting for atmospheric pressure, the RMSE dropped even lower to 2.7 ppm. Furthermore, including air temperature and water vapor mixing ratio resulted in a RMSE of 2.7 ppm and 2.1 ppm respectively. It is important to note that the temperature regression did slightly reduce the RMSE, but not significantly enough to be resolved with only two significant figures. Therefore, using the successive regression method, the RMSE of the observed difference dropped from 6.9 ppm to 2.1 ppm, a reduction of the error by over a factor of three. Fig. 2.6 shows the results and scatter plots for each step of the correction for this K30; Fig. 2.7 shows a difference plot at each step for this same K30 unit. Similar results

were observed for each K30 sensor evaluated and a summary can be found in Table 2.1.

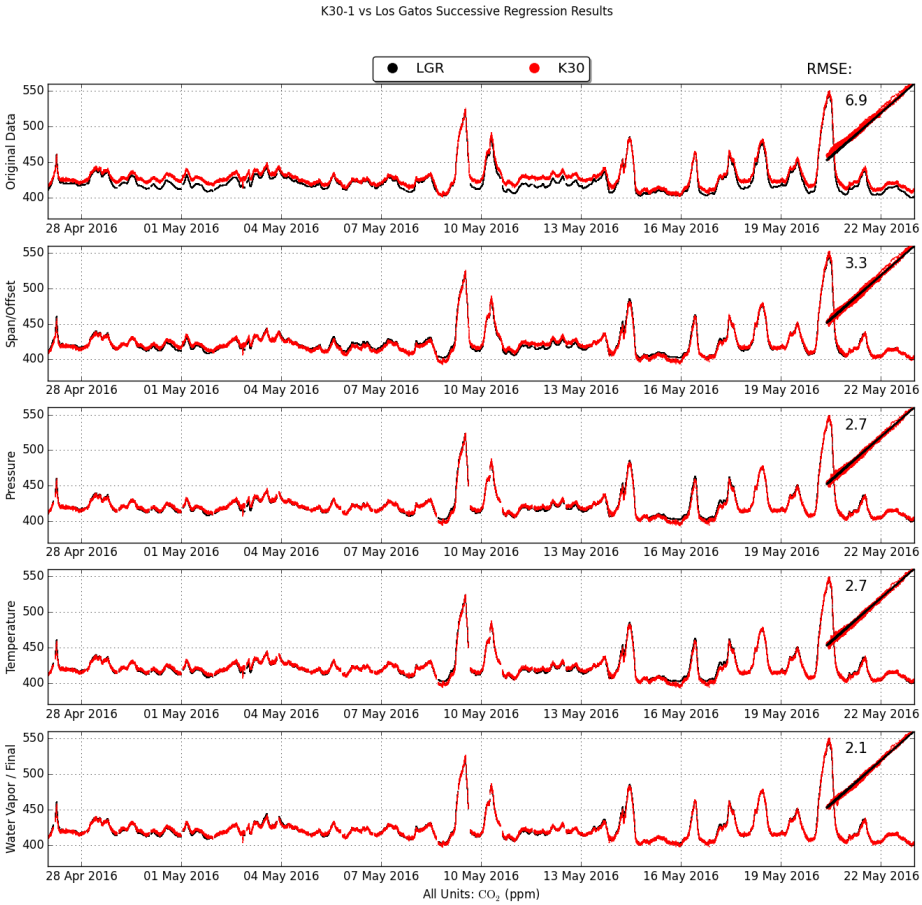

Figure 2.6. A continuous time series of 1-minute averages as well as scatter plots for K30 #1 compared to the LGR instrument during each step of the successive regression described in Sect. 2.5.1. Cumulative, in order from top to bottom: the original dataset, after correcting for span and offset, after correcting for pressure, after correcting for temperature, and finally, after correcting for water vapor. The root mean square error (RMSE) of the K30 data compared to the LGR at each step is annotated to the upper left of the scatter plot. This regression contains all data points observed in the evaluation period.

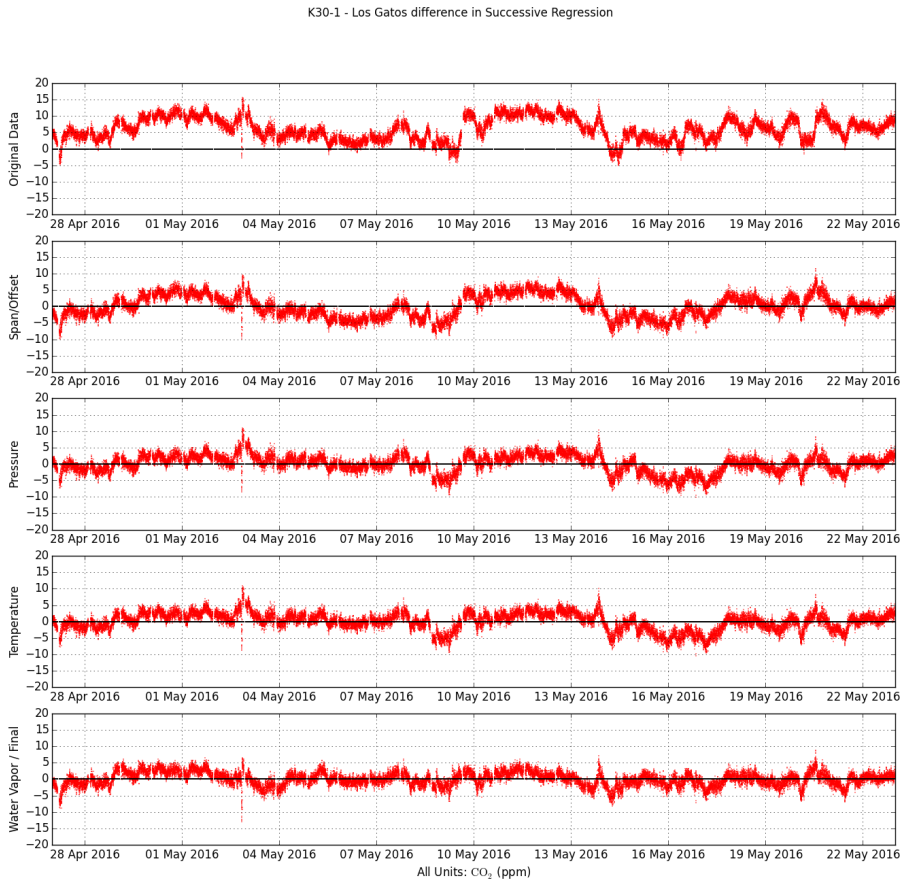

Figure 2.7. Difference plots for K30 #1 compared to the LGR during each step of the successive regression described in Sect. 2.5.1 and shown in Fig. 2.6 for 1-minute averages. Cumulative, in order from top to bottom: the original dataset, after correcting for span and offset, after correcting for pressure, after correcting for temperature, and finally, after correcting for water vapor.

2.5.2 Multivariate Linear Regression Method

Alternatively, a multivariate linear regression statistical method can be used to calculate the regression coefficients for each K30 sensor. This results in five correction coefficients a_n and b where n represents each independent variable, the dry CO_2 from the LGR, pressure P, temperature T, and water vapor mixing ratio q. Like the successive method above, these coefficients can be used in the following equation

along with the original K30 data, y, and the environmental variables to predict the true CO_2 concentration observed.

$$y_{corrected} = \frac{y-b-(a_1x_1)-(a_2x_2)-\cdots-(a_nx_n)}{a_0} \quad (2.4)$$

Using the multivariate regression function provided by Python-SciPy-Stats (Jones et al., 2001), differences from the LGR of the same K30 described in Sect. 5.1 were reduced to an RMSE of 2.1 ppm, slightly better than the iterative method. This consistently better performance from the multivariate method is shown in the other K30 sensors evaluated. Figure 2.8 shows the final results of the multivariate regression for the same K30 as in Fig. 2.6 and Fig. 2.7, as well as the difference between the corrected K30 dataset and the LGR. As with the univariate method, similar results were observed from each K30 sensor evaluated and a summary can also be found in Table 2.1.

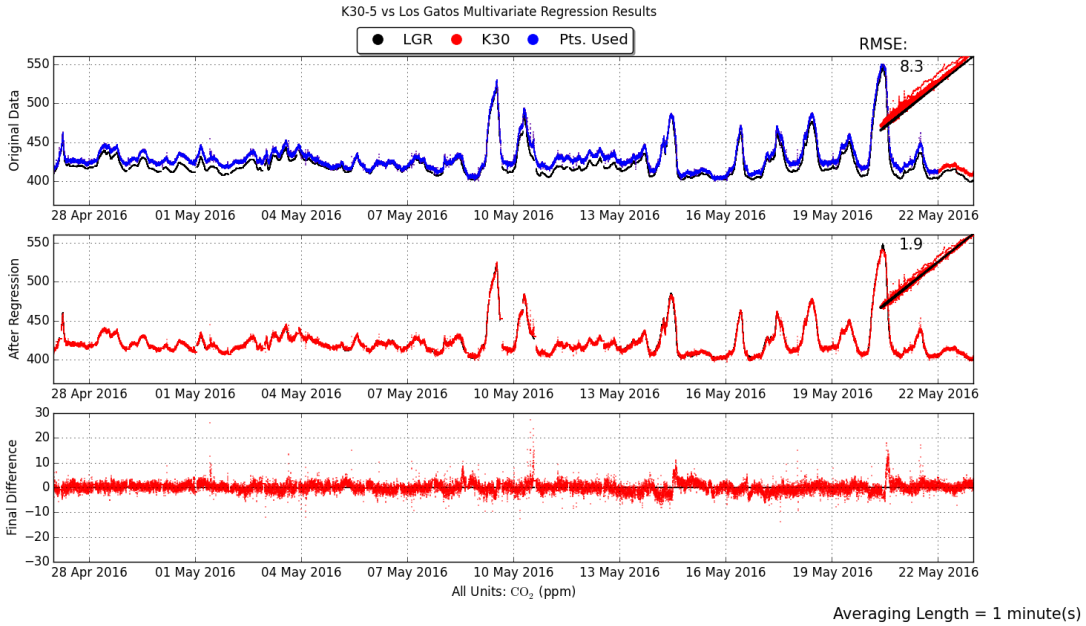

Figure 2.8. A continuous time series of 1-minute averages as well as scatter plots for K30 #1 compared to the LGR for the multivariate regression described in Sect. 2.5.2. Top panel: the original data, middle panel: final time series after correction, and the bottom panel: difference plot between the corrected K30 dataset and the original LGR dataset. The root mean square error (RMSE) of the K30 data compared to the LGR before and after the regression is annotated to the upper left of the scatter plot.

2.6 Discussion

2.6.1 Time Averaging

There are two observations to note based on the evaluation and analysis. First, both before and after the multivariate regressions, there are frequent shifts in the sign of the difference between each K30 and the LGR; these sudden changes occur at or around sunrise most days. Because of the rapid change in atmospheric CO_2 concentration at this time, the ambient calibration chamber may not be well mixed during this time period. Each K30 is located in a slightly different location in the ambient calibration chamber, and are all approximately 1 to 2 meters away from the LGR inlet. This effect, combined with the different response time of the K30s compared to the LGR, can lead to dramatic differences between what each K30

observes and what the LGR observes at the same timestamp for a short period of time each day.

Atmospheric inversion methods often use hourly averaged data from tower observations (McKain et al., 2012; Bréon et al., 2015; Lauvaux et al., 2016), so after the multivariate regression was applied, the K30 and LGR datasets were further averaged to 10 minute and hourly datasets. The average RMSE for the six K30s with the one-minute data is 2.3 ppm, 2.0 ppm for 10-minute averages, and 1.8 ppm for hourly-averaged data. Throughout this analysis period, one of the six K30s evaluated performed consistently worse than the others, and after removing it from the averages, the RMSE values dropped to 1.9 ppm, 1.6 ppm, and 1.5 ppm, for 1-minute, 10-minute, and hourly averages, respectively. Thus, by using hourly averages and discarding underperforming sensors, the average RMSE of the difference between the LGR and a K30 NDIR sensor can be reduced to approximately 1.5 ppm.

2.6.2 Regression Period

The RMSE described above and in Table 2.1 are for regressions calculated over the entire experiment period of approximately four weeks. One goal of this work is to develop a methodology to evaluate individual sensors quickly so that they can be used in scientific applications. In Fig. 2.9 the average RMSE calculated over the entire month of all six K30s is plotted with respect to the number of days used in the multivariate regression from Sect. 2.5.2. While the RMSE is generally minimized with increasing regression length, after a regression period of just a few days, the RMSE drops significantly from its initial values. Once a few diurnal cycles of varying

amplitude have been incorporated, as well as the synoptic scale variations in the atmosphere (with a time scale of around one week), the regression stabilizes. Thus, a regression length of around two weeks is recommended to maximize correction while minimizing the required amount of time the sensor needs to run concurrently with the LGR.

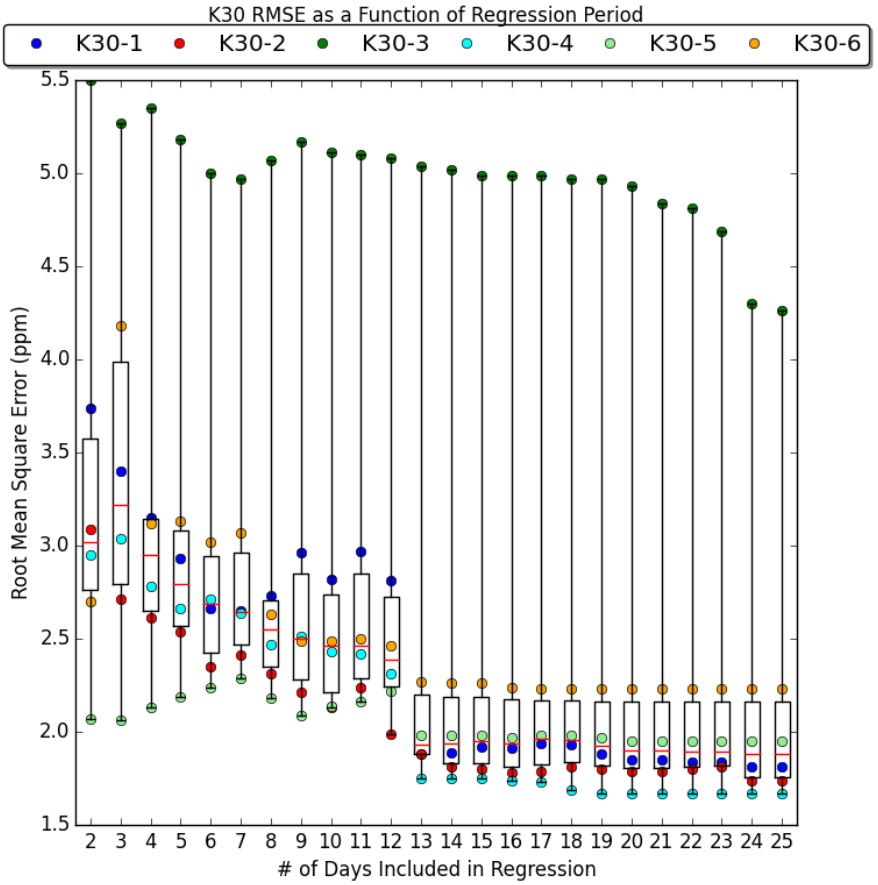

Figure 2.9. The RMSE of all six K30 NDIR sensors when compared to the LGR over the entire experiment as a function of how many days the regression analysis was performed. The colored dots represent each K30's RMSE, and the box plot shows the median in red, the first and third quartiles within the box, and the min and max values on the whiskers.

In Fig. 2.10, a multivariate regression is applied to the same K30 as described in the aforementioned sections and shown in Figs. 2.6, 2.7 and 2.8, but the

coefficients are calculated using only data from the first 15 days. The change in the RMSE between the two regressions is 0.1 ppm, going from 1.8 ppm when using all data points to 1.9 ppm when using only approximately the first half. This small, but not insignificant change is most likely attributed to the fact that during the first half of the evaluation period, the ambient CO_2 concentrations do not vary significantly, especially relative to the second half, where both the minimum and maximum values occur. In fact, when instead regressing for the last 15 days of the period, the RMSE is 1.8 ppm, a difference not distinguishable with only one decimal place. So as stated above, the diurnal cycles act as a range of calibration points, but values above and below what is included in the regression period may cause the corrected data to still have large errors during these periods, increasing the RMSE for the entire evaluation cycle. Based on these results, it is reasonable to assume that there is either no noticeable baseline drift or that it is assumed to be linear and removed by the multivariate regression in the sensors observed on the weekly to monthly timescales. The longer-term drift of the sensors for periods greater than one month is not known at this time, however, and would require a longer evaluation period of at least six months.

2.6.3 Generalized Regression Coefficients

All of the final RMSEs calculated in this analysis are from using individual regression coefficients for each K30 sensor. However, it would be beneficial to determine if a generalized set of regression coefficients could be applied to any K30 sensor, and what the RMSE over the evaluation period would be. To calculate the

generalized coefficients, the four slopes for each variable as well as the intercepts for each of the five remaining sensors were averaged together, K30-3 was omitted due to the fact that it was the poorest performing sensor, and that its coefficients were significantly different from the other five. After correction using the same set of coefficients, the RMSEs of the six sensors ranged from 3.1 ppm to as high as 23.9 ppm. The final RMSEs in some cases were higher than with the original, uncorrected data. Similar results were observed when the multivariate regression coefficients were calculated using the mean concentration of the five sensors. Thus, it appears that for each K30 sensor, an independent evaluation must be completed to provide observations with a sufficient level of quality.

2.7 Conclusions

The K30 is a small, low-cost NDIR CO_2 sensor designed for industrial OEM applications. Each of the sensors tested falls within the manufacturer's stated accuracy range of ±30 ppm ±3 % of the reading when compared to a high-precision CEAS analyzer, but these ranges are not particularly useful for scientific applications aimed at measuring ambient atmospheric CO_2. If these sensors are individually calibrated, selected for stability, and corrected for sensitivity to temperature, pressure, and RH, the practical error of these sensors is < 5 ppm, or approximately 1 % of the observed value. The final RMSE of the six K30 ranged between 1.7 ppm and 4.3 ppm for 60 s averaging times. Averaging for 200 s further reduces the noise by about 30 %, but longer times did not further improve precision. With errors in this range, these instruments could be used in a variety of scientific applications, including

observations at high spatial density to better represent the range and distribution of an urban or natural region's CO_2 concentration.

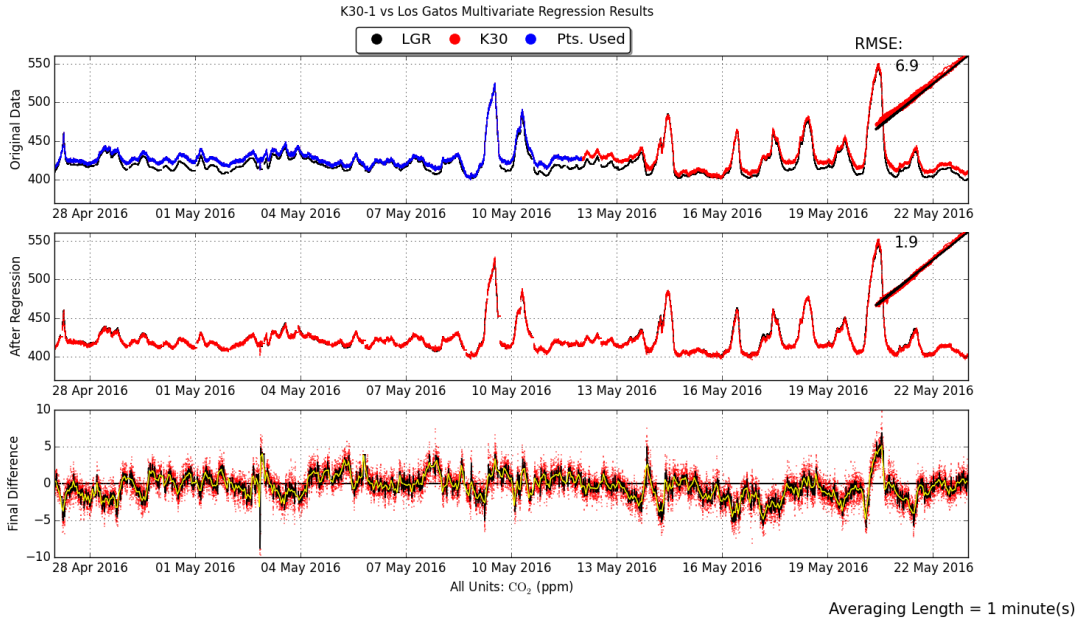

Figure 2.10. As depicted in Fig. 2.8, a continuous time series as well as scatter plots for K30 #1 compared to the LGR for the multivariate regression described in Sect. 5.2. Top panel: the original data, middle panel: final time series after correction, and the bottom panel: difference plot between the corrected K30 dataset and the original LGR dataset. However, this regression only includes the first 15 days of data (regression training data in blue, the entire dataset in red) to compute the correction coefficients. The difference plot (bottom) also shows running means for 10 minute (black) and hourly (yellow) averages.

Chapter 3 Investigating Sources of Variability and Error in Simulations of Carbon Dioxide in an Urban Region

Chapter based on manuscript under review for publication to Atmospheric Environment, August 2018.

Cory R. Martin[a], N. Zeng[a,b], A. Karion[c], K. Mueller[c], S. Ghosh[d], I. Lopez-Coto[d], K. R. Gurney[e], T. Oda[f,g], K. Prasad[d], Y. Liu[a*], R. R. Dickerson[a,b], and J. Whetstone[c]

[a] Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA
[b] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[c] Special Programs Office, National Institute of Standards and Technology, Gaithersburg, MD, USA
[d] Fire Research Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
[e] School of Life Sciences, Arizona State University, Tempe, AZ, USA
[f] Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[g] Goddard Earth Sciences Research and Technology, Universities Space Research Association, Columbia, MD, USA
*Now at: College of Geosciences, Texas A&M University, College Station, TX, USA

3.1 Introduction

As mentioned in Chapter 1, one of the most recent urban GHG campaigns is the National Institute of Standards and Technology (NIST) Northeast Corridor (NEC-B/W), currently encompassing the Baltimore, Maryland and Washington, DC metropolitan areas (Lopez-Coto et al., 2017; Mueller et al., 2018). This GHG observation network was implemented to demonstrate and improve measurement capabilities for quantifying anthropogenic GHG emissions from urban areas that cannot easily be disentangled from one another. It is expected that meteorological conditions are not spatially uniform across the region and are temporally impacted by distinct synoptic events. Although this campaign will ultimately consist of a sixteen-tower network, in 2016 only three towers were operational. The observations from

these three towers provide an opportunity to (1) investigate the ability to predict mole fractions using an atmospheric transport and dispersion model, along with prior flux distributions and (2) assess the relative impact of transport and prior errors on the simulated observations.

To achieve these objectives, an Eulerian transport model is employed that includes passive chemical tracers that use emission inventories as the surface flux along with initial and boundary conditions to generate 4D fields of atmospheric CO_2. In this way, we can vary the tracers to examine the sensitivity of the predicted mole fractions compared to the assumed meteorology. Eulerian models advect and disperse GHGs forward in time compared to Lagrangian approaches that use particle dispersion models operating backward from an observational 4D location. These are analogous approaches, but we employ a Eulerian model so that we can examine simulated meteorology for the entire domain to help us interpret model performance. In addition, we focus on CO_2 given the availability of data, specifically inventory data so that multiple inventories can be used to estimate errors resulting from emissions inventories. Simulated CO_2 mole fractions are compared to CO_2 observations from four in situ towers sites (three urban and one rural) in the NEC-B/W for the month of February 2016.

This chapter is outlined as follows: Section 3.2 describes the methods used for this analysis, including the model domain and configuration, the observation datasets used in the evaluation, and the emissions inventories as input to the transport model, Section 3.3 presents the observed and modeled CO_2 time series at specific locations, and compares the model meteorology and predicted mole fractions to observed

values. A discussion of the results is featured in Section 3.4 and a summary of results and conclusions are presented in Section 3.5.

3.2 Methods

The main component of the modeling framework described in this study is the Weather Research and Forecasting model coupled with chemistry (WRF-Chem), a non-hydrostatic, compressible model that provides passive tracer transport online with mesoscale meteorology forecasting capabilities (Grell et al., 2005; Skamarock et al., 2008; Beck et al., 2011). WRF-Chem has been modified to allow for separate passive CO_2 tracers for four anthropogenic emissions inventories. To evaluate the modeled CO_2's sensitivity to the tracer input, we employ inventories that are commonly used as prior anthropogenic fluxes in inverse modeling studies. In this paper, we refer to a tracer as a 4D mole fraction field of CO_2 whereas the emissions inventory refers to the 3D (or 2D if it does not have temporal variability) flux field. Additionally, a tracer for the biogenic component of the CO_2 concentrations is also included in this modified version of WRF-Chem since the mole fractions observed at tower locations are the integrated signal of both biospheric and anthropogenic fluxes on top of the global atmospheric concentration. For the subsequent analysis presented in this chapter, WRF-Chem was run for the month of February 2016. The month of February is used because it is assumed that anthropogenic emissions dominate the integrated atmospheric signals as observed from these tower locations during winter months, as the biosphere is assumed to be relatively inactive. Thus, the impact of

differences between various prior anthropogenic flux estimates can be better ascertained in the simulated observations.

3.2.1 CO_2 Observations

The NEC-B/W will ultimately feature a network of 16 observation sites (12 urban/suburban sites and 4 rural sites) to measure CO_2 and methane (CH_4) continuously. Operated by Earth Networks (EN), each site will feature a high-precision CRDS greenhouse gas analyzer and a calibration and data processing system similar to the in situ sites in the Los Angeles Megacities project (Verhulst et al., 2017). Additionally, a software-controlled valve system switches between multiple inlets, where available, to pull ambient air to sample from different heights above ground level. Data are quality controlled and averaged to hourly mole fractions reported on the WMO X2007 scale (CO_2) and X2004A scale (CH_4) for each inlet height.

In February 2016, three GHG observation tower sites had been established and were collecting continuous in situ CO_2 mole fraction measurements. The three sites are: HAL in Halethorpe, MD southwest of Baltimore (39.2552N, 76.6753W), NDC in the Tenleytown neighborhood of northwest Washington, DC (38.9499N, 77.0796W), and ARL in Arlington, VA (38.8917N, 77.1317W). Additionally, the NOAA / University of Virginia CO_2 observation site in Shenandoah National Park (SNP, 38.6170N, 78.3500W; Lee et al., 2012; Andrews et al., 2014; CarbonTracker Team, 2017) was used as a rural comparison site, as it is located at 1008 m above sea level putting it frequently above the local planetary boundary layer (PBL) at night

(Poulida et al., 1991). This site is also over 10 km from the nearest town, over 25 km from the nearest major highway, and far from most local anthropogenic influences. This analysis uses observations from all inlet heights when available, but for plotting purposes only the lowest inlet is shown for time series. These four site locations are shown in Fig. 3.1, with additional information in Table 3.1.

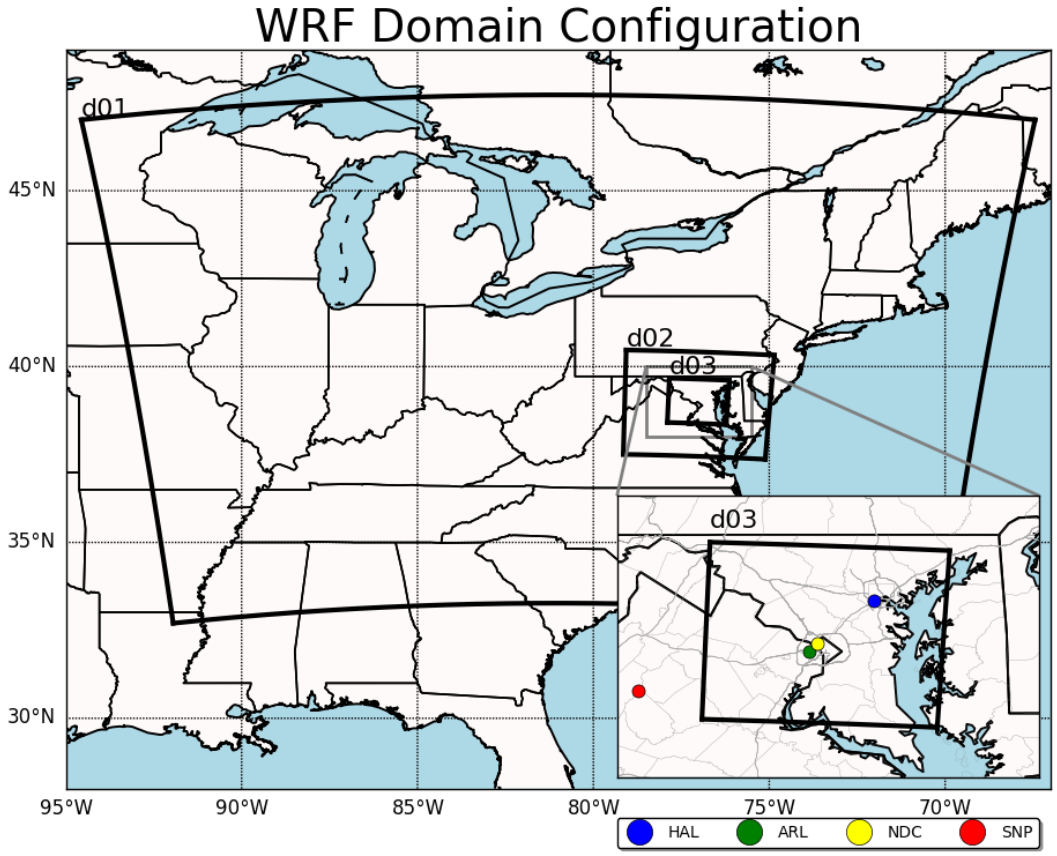

Figure 3.1. Map showing the WRF-Chem domain configuration used in this analysis. Domain d01 is modeled with 9km horizontal resolution, d02 with 3km, and d03 with a 1km horizontal resolution. The lower right inset shows the immediate area around d03 and the locations of the observing sites used: Shenandoah National Park (SNP; red circle), Arlington, VA (ARL; green circle), Northwest Washington, DC (NDC; yellow circle), and Halethorpe, MD (HAL; blue circle). Major highways are plotted as dark gray lines on the inset map along with the county boundaries in light gray.

Table 3.1. Summary of the four observation sites used in this study.

Site	SNP	ARL	NDC	HAL
Site Name	Shenandoah National Park	Arlington, VA	Northwest Washington, DC	Halethorpe, MD
Latitude (ºN)	38.6170	38.8917	38.9499	39.2552
Longitude (ºW)	78.3500	77.1317	77.0796	76.6753
Inlet Heights (meters above ground level)	17 m	50 m, 92 m	45 m, 91 m	29 m, 58 m
Site Elevation (meters above sea level)	1008 m	111 m	128 m	70 m
Data provider	NOAA/UVA	NIST/EN	NIST/EN	NIST/EN

3.2.2 WRF-Chem

A triply nested grid was defined for the WRF-Chem model configuration (Fig. 3.1). The outermost domain (d01) covers roughly the northeastern quadrant of the United States at a horizontal resolution of 9 km. The d01 extent was chosen because the predominant wind direction for the NEC-B/W is from the North and Northwest (Whelpdale et al., 1984) in February, and this extent generally captures the incoming CO_2 from areas as far away as Chicago, IL. Within this parent domain is an intermediate two-way nested domain (d02) with a resolution of 3 km. An additional fine-scale domain (d03) is nested within d02; it features a horizontal resolution of 1 km that covers the metropolitan areas of the NEC-B/W. A description of the parameterizations and options used for each WRF-Chem domain is provided in Table 3.2 (Chou et al., 2001; Hong et al., 2004,2006; Kain 2004; Mlawer et al., 1997; Tewari et al., 2004).

Table 3.2. Summary of the three WRF domains and their configurations

Domain	**d01**	**d02**	**d03**
Horizontal Resolution	9 km	3 km	1 km
Vertical Levels	50 (from surface to 50 hPa)		
Microphysics	Single-Moment 5-class		
Radiation	RRTM longwave and Goddard shortwave		
Cumulus Parameterization	Kain-Fritsch	Kain-Fritsch	None
PBL Scheme	Yonsei University		
Land Surface	Noah Land Surface Model		

Meteorological initial and boundary conditions are provided by the National Oceanographic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR), a product with a horizontal resolution of 32 km, 30 vertical layers, and three-hourly output (Mesinger et al., 2006). Because the simulation runs for the entire month, sea surface temperatures are also included as boundary conditions in the WRF-Chem model. The high-resolution version of the NOAA NCEP real-time, global, sea surface temperature analysis (RTG_SST_HR) with a horizontal resolution of 1/12 degree and daily output is used (Thiébaux et al., 2003; Gemmil et al., 2007).

Initial and boundary conditions for the background CO_2 concentrations are provided by NOAA Earth System Research Laboratory's (ESRL) CarbonTracker Near Real-Time gridded product (Peters et al., 2007; https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT-NRT/index.php). This is a 3D mole fraction product with three-hourly output and a horizontal resolution of 1º over North America. This background value is available as a separate tracer at all hours of the simulation, and is added to the other tracers for the total predicted CO_2. All tracers

resulting from the anthropogenic emissions inventories have initial and boundary conditions of zero ppm. The model-simulated CO_2 mole fraction determined at a point in time and space is the sum of the tracer associated with the specified anthropogenic emissions inventory, the biospheric flux tracer, and the advected background CarbonTracker mole fraction.

Ten Automated Surface Observing System (ASOS) or Automated Weather Observing System (AWOS) surface observation stations are used to evaluate WRF's prediction of near-surface wind speed and direction. The datasets were downloaded from the National Weather Service's Meteorological Assimilation Data Ingest System (MADIS) and processed by the supplied API. The observations for wind are generally sited at 10m above ground level, and the WRF interpolated 10m winds are used for the comparison. Both the observations and the model output provide the U and V components of the wind vector, which are converted to speed and degrees from north for subsequent comparison.

To evaluate WRF's simulation of the vertical temperature profile of the atmosphere as well as its calculation of the PBL, observations from the Aircraft Communications Addressing and Reporting System (ACARS) are used. Available from the MADIS API, profiles are created when certain aircraft take off and land at airports. Two of the three airports (KBWI, Baltimore-Washington Thurgood Marshall International and KDCA, Reagan National) located within the NEC-B/W regions have sufficient data available for February 2016. The WRF PBL scheme calculates the height of the PBL as an output variable, but the ACARS dataset only provides vertical profiles of observed quantities such as temperature and pressure. To compare,

an algorithm is used to estimate the observed and modeled PBL height by computing the vertical potential temperature gradient and finding where the gradient is at its relative maximum. For each hour, all available ACARS profiles meeting the following criteria are used: the lowest reported data point must be below 300 m above ground level, there must be at least 10 data points, and the computed PBL height must be below 2100 m (in case the tropopause is detected by the algorithm). The PBL height is computed from all valid profiles, and then the mean of these computed heights is used. Any hour containing fewer than three individual profiles is not used. To supplement the ACARS profile data at the third airport (KIAD; Washington-Dulles International), the radiosonde profiles launched by the National Weather Service in Sterling, VA (KLWX) are used. The office, located onsite at the airport, launches radiosondes typically twice per day (0000 UTC and 1200 UTC; 7pm and 7am local time), so the temporal coverage of observed profiles is sparse compared to the other two airports.

For all hours, WRF has a positive wind speed bias at all but one site (KDCA, Reagan National Airport) with an average over all ten locations of 1.2 m/s. This slight positive bias in wind speed is consistent with previous comparisons of WRF to observations in similar work (e.g. Nehrkorn et al., 2012; Feng et al., 2016). The average wind direction bias over all sites and for the entire month is approximately 2.8º. The standard deviation of the average difference between model and observations is virtually the same as the mean bias for wind speed, with an average standard deviation of 1.2 m/s to one decimal place.

When averaged over the entire evaluation period, the comparisons between estimated PBL heights from WRF and the observations at each airport exhibit different behavior. Both the computed PBL height from the potential temperature gradient and the YSU predicted PBL height (using the Richardson number) are compared to the calculated height from the observed profiles. KBWI is the best performer, with a mean error of the computed height of 11 m, compared to an error of -145 m for the YSU predicted height. At KDCA, the YSU predicted height is 30 m too high on average, whereas using the potential temperature profile results in a higher bias of 325 m. Finally, at KLWX, using only the radiosondes, the mean difference over the month is negative for both methods, with an average of -118 m below observations for the computed height and -481 m for the YSU provided height. For most days, KLWX only has profiles twice a day, and as such may not be representative of the model's overall performance. Overall, the YSU estimated PBL height is an average of -71 m from the observations, and the potential temperature profile method results in a 150 m high bias. The aforementioned meteorological fields from WRF agree well with observations on average, but can vary greatly for any given hour. See Fig. 3.2 for time series of the mean (solid line) and range (min and max; shaded regions) difference between WRF and observations for all three variables: PBL height (blue is potential temperature; red YSU), wind speed, and wind direction.

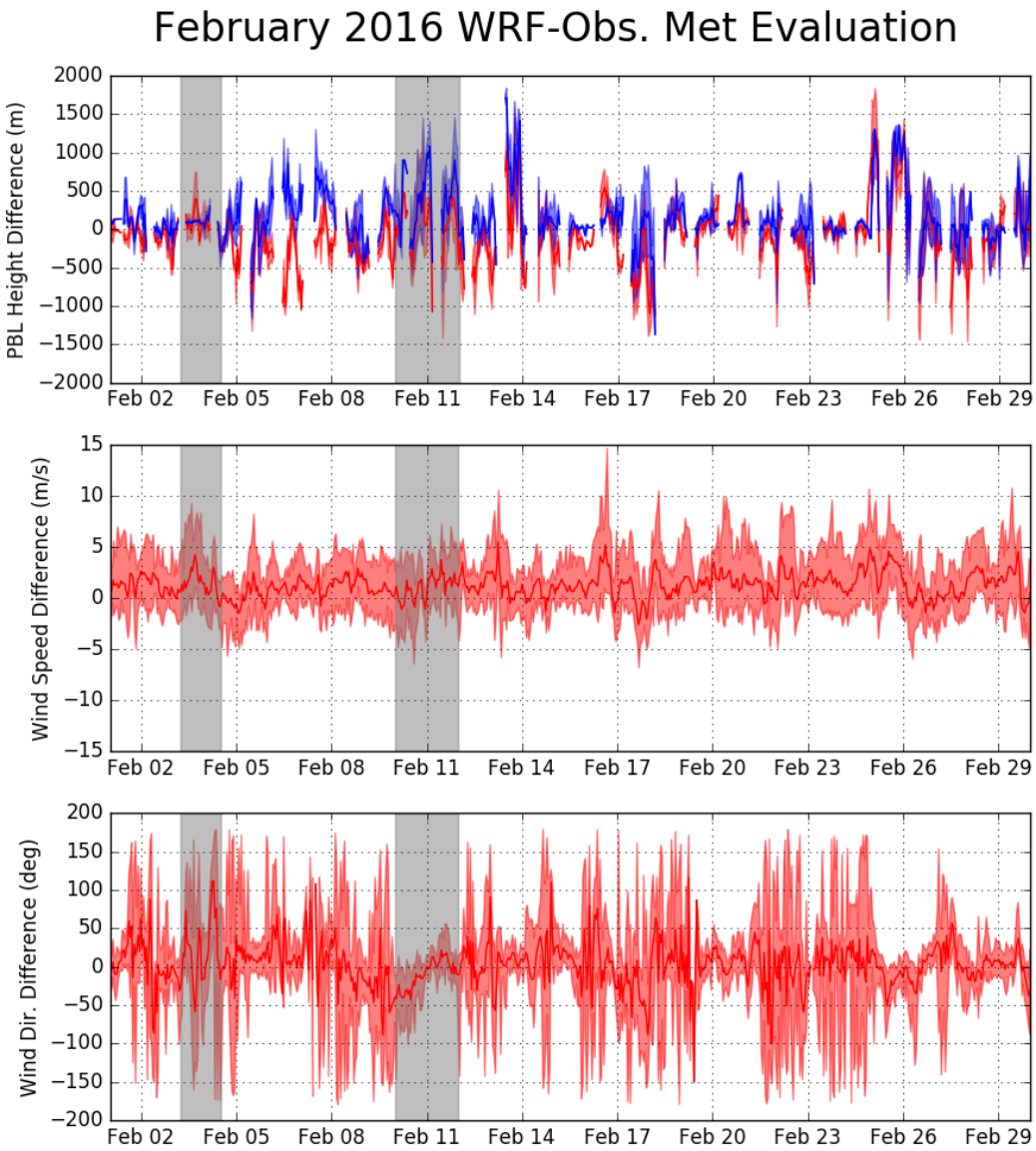

Figure 3.2. Mean (solid line) and spread (shaded area between minimum and maximum values) of the difference between WRF and observations for: PBL height (YSU computed PBL height in red; potential temperature profile computed in blue), wind speed, and wind direction. For PBL, observations are for three airports (KDCA, KBWI, KIAD), and for wind observations, ten sites in domain d03 are used. Positive values indicate a larger quantity from the model. The two cases are shaded in gray, as in Figs. 3.4 and 3.5.

3.2.3 Emissions Inventories

To evaluate whether the modeled observations are impacted more by the underlying emissions or transport, an ensemble of tracers of atmospheric CO_2 resulting from different emissions inventories are used within the same transport model simulation using WRF-Chem. For this study, four different anthropogenic CO_2 emissions inventories are used: EDGAR, FFDAS, ODIAC, and Vulcan. The following paragraphs and Table 3.3 provide details of each inventory.

The first inventory employed is EDGAR, i.e. the Emissions Dataset for Global Atmospheric Research version 4.2 (Olivier et al., 2005; http://edgar.jrc.ec.europa.eu). EDGAR is a global emissions product with a horizontal resolution of 0.1º, and provides average fluxes for the year 2010 based on the International Energy Agency's (IEA) energy budget statistics (IEA, 2012). The emissions are then distributed on the 0.1º x 0.1º grid by incorporating population density, road networks, and the locations of point sources and industrial processes.

The Fossil Fuel Data Assimilation System (FFDAS; Rayner et al., 2010; Asefi-Najafabady et al., 2014) is also used. As with EDGAR, FFDAS is a global product with a horizontal grid of 0.1º x 0.1º but unlike EDGAR, it features hourly varying anthropogenic fluxes for the entirety of 2015. FFDAS utilizes the Kaya Identity, a method to estimate emissions based on economic factors, as well as information on national fossil fuel CO_2 emissions, satellite-derived nightlights, population density, and power plant information to estimate flux at each grid point.

The Open-source Data Inventory for Anthropogenic CO_2 (ODIAC; Oda and Maksyutov, 2011; Oda and Maksyutov, 2015; Oda et al., 2018) is the third inventory

used in WRF-Chem. It is the only dataset of the four chosen with a finer horizontal resolution of approximately 1 km, or ~0.01 º. Using the total emissions estimated by the Carbon Dioxide Information and Analysis Center (CDIAC) at the US Department of Energy's Oak Ridge National Laboratory, the locations of point sources and satellite-derived nightlights are then used to distribute the emissions onto the 1 km grid. Monthly total fluxes are provided by ODIAC for each month projected using statistical data from the energy company BP with the most recent version for the year 2015.

The fourth inventory is Vulcan (Gurney et al., 2009) which is a 10 km x 10 km fossil fuel emissions dataset for the United States for the year 2002. Unlike the others, the emissions of individual buildings, power plants, roadways, and other sectors are each characterized, and then aggregated to this 10 km x10 km grid. This provides a higher level of detail both spatially and temporally, but with the limitations of being much older than the other inventories, and only covers the coterminous United States, so parts of domain d01 have no emissions in this tracer.

Table 3.3. Summary of the four anthropogenic CO_2 emissions inventories used within this study. For ODIAC, both the domain average sum is provided with and without temporal scaling added to the dataset.

Inventory	EDGAR	FFDAS	ODIAC	Vulcan
Version	4.2	2.2	2016	2.2
Horizontal Resolution	0.1 degree	0.1 degree	0.01 degree	10 km
Created for Year	2010	2015	2015	2002
Temporal Resolution Provided	Yearly	Hourly	Monthly	Hourly
Domain d01 Average Sum (Tonnes C / hr)	88416	104709	97732 106231 * (no TIMES scaling)	97469
Domain d03 Average Sum (Tonnes C / hr)	2112	3622	2332 2534 * (no TIMES scaling)	2882
Citation	Olivier et al., 2005	Rayner et al., 2010; Asefi-Najafabady et al., 2014;	Oda and Maksyutov, 2011; Oda et al. 2018	Gurney et al., 2009
Available at:	http://edgar.jrc.ec.europa.eu	http://www.gurneylab.org/portfolio-item/ffdas/	http://db.cger.nies.go.jp/dataset/ODIAC/	http://vulcan.project.asu.edu

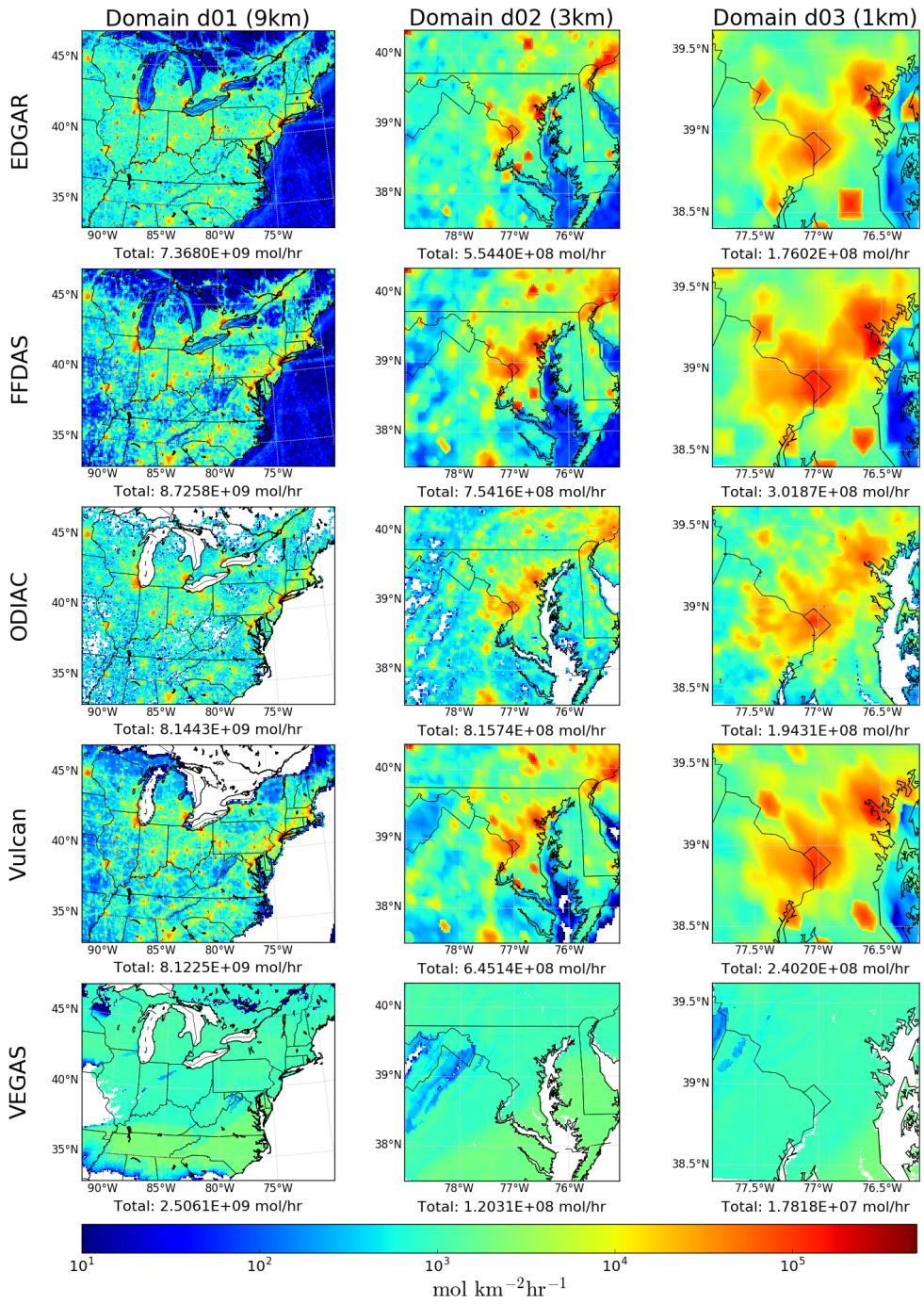

Figure 3.3. Average CO_2 hourly fluxes for the four emissions inventories and the VEGAS biospheric model for all three WRF-Chem domains in February 2016. Non-positive values (zero and negative) are shown as white. The hourly average flux for the month for each dataset summed over the entire domain, is shown below each map.

The inventories are interpolated in time and space to ensure consistency. Each one is linearly interpolated from its native grid to the three WRF-Chem domains. Additionally, the Temporal Improvements for Modeling Emissions by Scaling (TIMES; Nassar et al., 2013) scale factors are applied to ODIAC and EDGAR to provide weekly and diurnal variations to these two inventories. However, we also use the native monthly ODIAC product as input to WRF-Chem (aka ODIACFIX) as it allows us to investigate the impact of diurnal and weekly varying fluxes on simulated observations. Note that TIMES scaling results in an approximate decrease of emissions of 8.5% when averaged over the entire month because while the daily average remains the same for weekdays, the scaling factor causes a reduction for weekend hours and the number of each day of the week is not the same in any given month. The impact of the TIMES scaling on the simulated observations will be discussed in Section 3.3.1.

We further ensure consistency between the inventories by shifting the inventories so that the calendar days and hours are the same across all emission products. For example, the fluxes for February 2, 2015 of FFDAS are used for February 1, 2016 as they are both Mondays. In addition, since the inventories were generated for a year differing from the modeled year, ratios are used to scale each emission product using national totals from the U.S. Energy Information Administration's (EIA) Monthly Energy Review (https://www.eia.gov/totalenergy/data/monthly/) for each day of February as shown in Equation 3.1. Emission products will be referred to as tracers from henceforward to

correspond to WRF-Chem nomenclature. Fig. 3.3 shows a map of the hourly mean flux of CO_2 from each inventory interpolated to all three WRF-Chem domains.

$$Inventory_{2016} = Inventory_{Feb.Year} \times \frac{EIA\ Total_{Feb.2016}}{EIA\ Total_{Feb.Year}} \quad (3.1)$$

3.2.4 Vegetation Model

As mentioned earlier, a dynamical vegetation model has also been coupled to WRF-Chem to provide the contribution of biogenic fluxes to the simulated observations. The VEgetation-Global-Atmosphere-Soil (VEGAS) model (Zeng et al., 2005) is coupled offline with WRF-Chem to provide hourly biospheric CO_2 flux. Because VEGAS features carbon pools and dynamic vegetation growth, the model must first be spun up on the domain to achieve a climatology. For this analysis, VEGAS is first initialized by forcing it with the Climate Forecast System Reanalysis version 2 (CFSRv2; Saha et al., 2014) calibration climatologies for the years 1981 to 2010. The CFSR climatology dataset is first regridded to the WRF-Chem domains, and then the model is run for 100 years using this calibration climatology repeatedly to reach equilibrium. To generate the land to atmosphere carbon flux, VEGAS uses the WRF-Chem meteorological output variables (2 m temperature, 2 m specific humidity, hourly precipitation, 10 m winds, skin temperature, and total net radiation) as well as the WRF domain topography, emissivity, and albedo. Figure 3.3 shows the hourly average biospheric flux from VEGAS on all three WRF-Chem domains in February 2016; including areas of net uptake (the white region in the south part of domain d01). The biospheric VEGAS tracer has been added to all the simulated CO_2

values shown although its average contribution to the simulated CO_2 mole fraction is less than 0.25% of the total over the month. Additionally, no evaluation has been performed on the biospheric fluxes from VEGAS at these spatial scales further highlighting why we focus on winter months for this analysis.

3.3 Results

In this section we assess WRF-Chem's ability to simulate the atmospheric CO_2 in the NEC-B/W by comparing modeled CO_2 mole fractions at four locations to high-accuracy in situ observations from the three urban and one rural tower sites. First, these datasets are compared over the entire month-long simulation to determine the overall performance of the model. We then select specific time periods of the month to diagnose possible causes of both high and low performing scenarios. These two analyses help us to evaluate the performance of WRF-Chem in modeling transport and dispersion of urban CO_2 and whether there is sufficient skill in the model for use with various GHG flux estimation methods.

3.3.1 Overall Model Performance

3.3.1.1 Tower Observations

Not surprisingly, the magnitudes and variability of the observations from the towers are different for the rural site compared to those from the urban towers (Fig. 3.4). Over the four locations, the lowest observed CO_2 of the four sites was typically at SNP (with an average of 412.2 ppm vs. the mean of the urban sites at just over 421 ppm) due to its rural location and high altitude, frequently above the PBL in the free troposphere. The amplitude of the observational diurnal cycle at SNP is also smaller

than the urban towers but can vary from day to day depending on the synoptic weather situation. The variability in the diurnal cycle at the urban sites is much greater (frequently as high as 50 ppm, but occasionally under 5 ppm) which indicates that synoptic events have a large impact on urban CO_2 observations given the magnitude and variation of the underlying flux distribution in such areas.

3.3.1.2 Simulated Observations

In general, WRF-Chem generated mole fractions with similar magnitudes and variabilities to the observed mole fractions from the four tower sites. Consistent with the observations, the relative magnitudes of the simulated mole fractions at urban towers are larger than those from the rural site and they exhibit more diurnal variability. In addition, when looking at the model predicted mole fractions in Fig. 3.4 across all sites, one can note the variations in synoptic and diurnal cycles are similar to the observed time series. These results provide evidence that the model is able to reasonably recreate the time series of CO_2 mole fractions when looking at the complete time series, but an in-depth analysis is required to determine its performance for a particular day or period.

During certain unfavorable meteorological conditions, the spread of the individual emission tracers increases even though their overall variability remains proportionally the same. In terms of overall magnitudes, the differences in the minimum and maximum daily values (i.e. differences between the observed and modeled mole fractions) can be at times quite significant. These large differences correspond to synoptic scale weather patterns (3 d to 5 d) that also create the variability in the observations as discussed in Section 3.3.1.1. Depending on the day,

the differences between modeled and observed CO_2 mole fractions throughout the day can vary by an order of magnitude from less than 5 ppm to over 50 ppm. These synoptic weather conditions will be discussed further in Section 3.3.2.

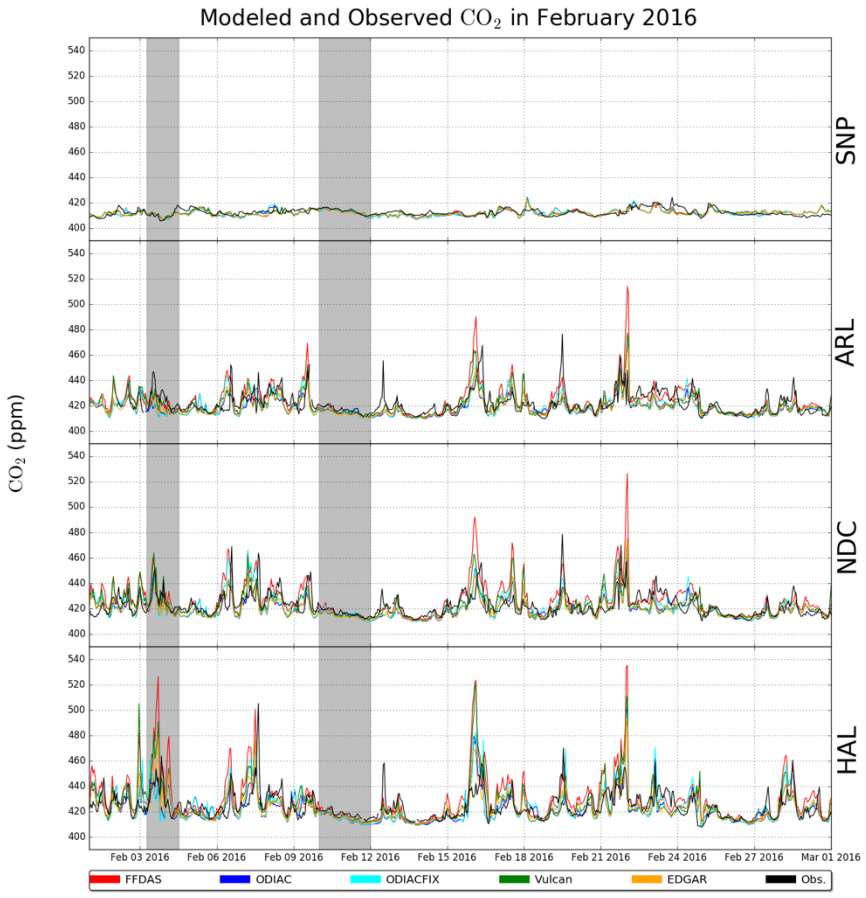

Figure 3.4. Time series of hourly averaged modeled versus observed CO_2 mole fractions at four observing sites for all hours of the day. The black lines are the observed values, and each color represents the model-simulated CO_2 interpolated to that location and inlet height (only the lowest inlet levels are plotted at the 3 urban sites). The model-simulated mole fraction at a point in time and space is the sum of an anthropogenic tracer generated from a specific inventory plus the VEGAS biospheric flux tracer plus the background CarbonTracker advected value. From top to bottom: Shenandoah National Park (SNP), Arlington, VA (ARL), Northwest Washington, DC (NDC), and Halethorpe, MD (HAL). FFDAS predicted values are in red, ODIAC in blue, ODIAC without temporal scaling in light blue, Vulcan in green, and EDGAR in orange. Gray shaded areas are scenarios described in detail in Section 3.3.2.

Table 3.4. Results of a linear regression performed at each observing site and for both inlets where applicable for all five tracers. For each regression, the observed dataset is used as the independent variable, x, and the model predicted values are used as the dependent variable, y. Provided in the table are the: bias before the regression (ppm), slope, intercept (ppm), the coefficient of determination (R^2), the root-mean-square error before the regression (RMSE; ppm), the number of outliers using the double median absolute deviation (out of 696 hours), and the mean absolute error before the regression (MAE; ppm), for each simulated time series dataset at each observing level.

Site Code	Tracer Source	Inlet	Bias	Slope	Intercept	R^2	RMSE	# outliers	MAE
ARL	FFDAS	50m	1.47	0.77	98.95	0.79	7.26	200	4.87
ARL	ODIAC	50m	-3.54	0.58	174.28	0.14	6.54	167	5.06
ARL	ODIACFIX	50m	-2.27	0.59	170.80	0.38	6.96	171	5.23
ARL	Vulcan	50m	-1.34	0.61	162.78	0.53	6.47	174	4.66
ARL	EDGAR	50m	-3.25	0.58	172.84	0.21	6.38	185	4.89
ARL	FFDAS	92m	-0.93	0.74	109.54	0.76	5.61	160	3.97
ARL	ODIAC	92m	-4.54	0.49	208.79	0.31	6.42	143	5.07
ARL	ODIACFIX	92m	-3.78	0.55	186.30	0.03	6.36	158	4.97
ARL	Vulcan	92m	-2.80	0.58	175.44	0.29	5.61	174	4.10
ARL	EDGAR	92m	-4.35	0.50	204.97	0.24	6.15	149	4.77
HAL	FFDAS	29m	5.25	0.88	56.62	0.42	11.43	181	7.56
HAL	ODIAC	29m	-2.40	0.65	146.57	0.48	8.24	190	6.19
HAL	ODIACFIX	29m	-0.52	0.81	78.91	0.87	9.06	195	6.67
HAL	Vulcan	29m	1.08	0.79	87.94	0.86	9.67	204	6.73
HAL	EDGAR	29m	-1.47	0.66	142.61	0.61	8.18	187	6.00
HAL	FFDAS	58m	4.52	0.90	44.30	0.18	8.98	174	6.15
HAL	ODIAC	58m	-1.34	0.67	137.63	0.63	7.12	171	5.34
HAL	ODIACFIX	58m	0.26	0.78	94.17	0.84	7.72	190	5.69
HAL	Vulcan	58m	1.23	0.77	97.68	0.80	7.99	201	5.54
HAL	EDGAR	58m	-1.04	0.64	152.35	0.61	6.58	178	4.86
NDC	FFDAS	45m	3.84	0.81	84.91	0.53	8.58	198	5.87
NDC	ODIAC	45m	-2.84	0.56	182.13	0.28	6.96	157	5.31
NDC	ODIACFIX	45m	-1.25	0.64	151.44	0.58	7.41	177	5.59
NDC	Vulcan	45m	0.00	0.69	129.66	0.73	7.16	190	5.18
NDC	EDGAR	45m	-2.83	0.58	173.72	0.30	6.84	188	5.24
NDC	FFDAS	91m	1.56	0.67	141.20	0.64	7.11	180	4.78
NDC	ODIAC	91m	-3.06	0.54	191.69	0.07	6.13	164	4.74
NDC	ODIACFIX	91m	-2.11	0.56	181.34	0.30	6.39	164	4.85
NDC	Vulcan	91m	-0.93	0.50	207.83	0.35	6.25	158	4.40
NDC	EDGAR	91m	-2.97	0.48	217.60	0.00	6.11	169	4.69
SNP	FFDAS	17m	-0.26	0.51	200.36	0.43	2.19	116	1.72
SNP	ODIAC	17m	-0.68	0.52	196.53	0.32	2.37	131	1.86
SNP	ODIACFIX	17m	-0.50	0.54	189.95	0.41	2.40	132	1.86
SNP	Vulcan	17m	-0.25	0.48	215.68	0.28	2.20	112	1.74
SNP	EDGAR	17m	-0.56	0.47	219.25	0.30	2.24	116	1.76

To characterize the performance of the simulated mole fractions using WRF-Chem relative to observations while accounting for the differences in emissions inventories, a linear regression was performed for all five modeled time series at each site (and both inlets where applicable) against their respective observed time series.

The results from these regressions are available in Table 3.4. For this regression, outliers, defined as differences larger than three standard deviations between the mean of both the observations and the WRF predicted values are removed. The resulting statistics help discern whether the simulated mole fractions, on average, deviate strongly from the observations, are linearly related, and reflect the true variability.

Overall, for all of the urban locations, FFDAS has the strongest liner relationship to the observations with a slope closest to unity. However, FFDAS has the largest R^2 for only four out of the seven timeseries (those from two inlet heights at each of the three urban towers and from SNP) for the February 2016 model simulation. However, the lowest R^2 value is associated with the modeled mole fractions using FFDAS at both the HAL inlet levels. This could potentially be due to the location of HAL near large FFDAS point sources in Baltimore that are redistributed onto the native WRF-Chem grid. For the other inventories, the slopes and R^2 varies across all towers.

The slopes associated with simulated observations from the other four anthropogenic tracers have a small range of spread between inventories, with the spread being between 0.03 and 0.16 depending on observing site. The slopes closest to zero tend to be either ODIAC or EDGAR depending on the observing point location. The fact that these two inventories have similar slopes is not totally unexpected, as their derived emissions may be distributed spatially in a similar manner, albeit at different resolutions, and the TIMES scaling factors are applied to both of them.

The other statistics indicate that the performance of the modeled time-series is dependent on local conditions (i.e. meteorology or flux distribution) around each tower site since no single tracer consistently out-performs the others. For example, the MAE associated with the five anthropogenic tracers varies across tower inlets for a single inventory (such as FFDAS from 3.97 ppm to 7.56 ppm or Vulcan from 4.40 ppm to 6.73 ppm). FFDAS generally has the highest MAE with EDGAR or Vulcan usually having the lowest, but again, there is no consensus on any best or worst performer at each observing site. This suggests that model performance should not be assessed on average across the entire domain but rather locally around tower sites.

After calculating these linear regressions, the fitted datasets can be subtracted from the original modeled time series to see where the linear fit is not valid. Figure 3.5 shows the residuals of each linear fit from the observed CO_2 at each site where the five colors represent the different anthropogenic tracers in WRF-Chem. As with the slope, FFDAS (red) has a consistently larger absolute residual value (5.1 ppm) than the other four datasets (ODIAC (blue): 3.6 ppm; ODIACFIX (cyan): 4.3 ppm; Vulcan (green): 4.3 ppm; EDGAR (orange) 3.5 ppm) for February 2016 across the observing sites, likely due to the periodic high values skewing the linear fit as noted earlier. The residual plot also shows clearly periods where the simulated CO_2 deviates greatly from the observations for all tracers. This suggests that at times 1) the synoptic scale background CO_2 provided by CarbonTracker may not be resolved correctly, 2) there are sufficient errors in the meteorological transport, or 3) VEGAS is under-predicting respiration during this period. Two of these cases (the dark gray shaded regions in Fig. 3.5) will be described in detail in Section 3.3.2.

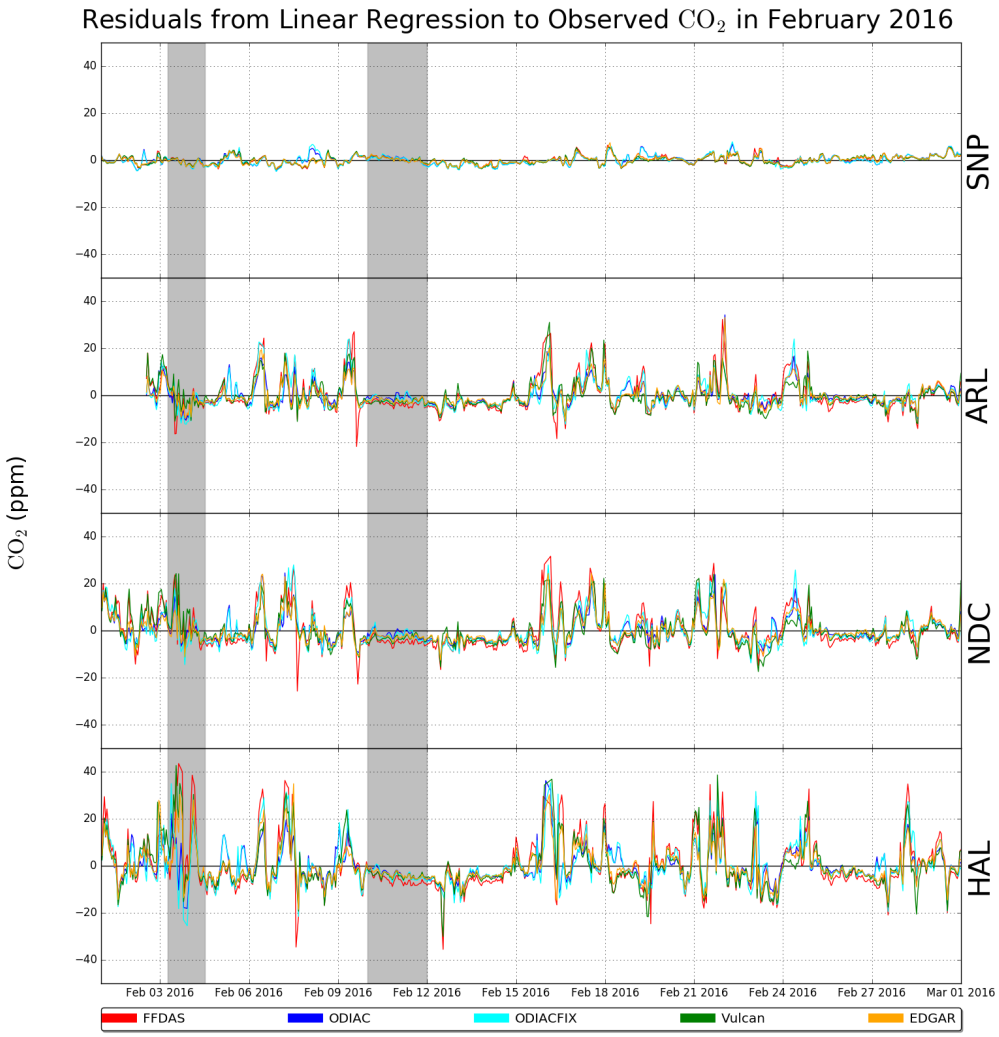

Figure 3.5. Residuals of a linear regression between the observed CO_2 and each tracer at all four observing sites (lowest inlet only at the 3 urban sites). See Section SI2 of the supplemental information for the regression equation used for this analysis. The different colors represent the five different tracers from the multiple emissions inputs. The dark gray shaded areas are scenarios described in detail in Section 3.3.2 which were also shown on Figure 3.4.

To investigate the impact of bias on the modeled vertical mixing and its representation of the PBL, the simulated mole fractions are analyzed using (1) all hours of the time series, and (2) afternoon hours only (12 pm to 4 pm local time; 17 UTC to 21 UTC), both without removing any outliers. Afternoon observations are typically used in inversions since it is assumed that the meteorological models can

better represent well mixed conditions typically found in the middle of the day. Figure 3.6 shows the monthly mean bias of simulated CO_2 mole fractions for all hours and afternoon hours only for all five tracers at the observing sites and different inlet heights.

The spread of the monthly bias from all tracers ranges from -4.5 ppm to 5.2 ppm for the five simulated observational timeseries encompassing all hours of the day. FFDAS has a slight positive mean bias at all three urban sites (the largest being at HAL and at NDC for the lowest inlet). This is consistent with the domain mean hourly averaged flux in domain d03 being the largest for FFDAS as shown in Fig. 3.3 and Table 3.3. Conversely, as expected from the flux summaries in Section 3.2.3, the most negative bias tends to be from either ODIAC or EDGAR as they are diurnally scaled using the TIMES dataset and have the lowest domain mean hourly averaged flux in the urban domain. The impact of the TIMES scaling is clearly demonstrated by the fact that ODIACFIX has a smaller bias than that of ODIAC.

When considering only afternoon hours, the spread in the monthly bias is smaller, ranging from (-3.9 ppm to 2.1 ppm). The FFDAS tracer yields the highest simulated CO_2 mole fraction (similar to all hours), and is the only inventory that has a clear positive bias. Although mostly negative, the Vulcan tracer has a near zero bias at both the HAL inlets. When looking at afternoon hours only, the ODIAC and ODIACFIX tracers are virtually the same as the TIMES scaling factors are based on a value of one for the mid-afternoon (the emissions are scaled down overnight). The mean bias from EDGAR during these periods is similar to that from ODIAC, with a negative mean bias of approximately 3 ppm to 4 ppm from the observations. This

range of biases (FFDAS positive to ODIAC and EDGAR being the lowest) is consistent with their respective rankings in the areal total anthropogenic flux for February 2016. Using afternoon hours only appears to reduce the spread of the modeled time series overall compared to the complete datasets but this is dominated by the large reduction in spread at HAL and the lower inlet at NDC.

Generally, WRF-Chem using these emissions inventories tends to underpredict near-surface CO_2 as shown in Fig. 3.6 (red and blue lines represent the average associated with all hours and afternoon hours respectively). On average, the mean of the five modeled time series is below the observed values, except for the all-hours datasets at HAL, likely largely the result of an underestimation of emissions. The smaller bias at SNP implies that a bias in the CarbonTracker background or the biospheric flux may also contribute to the low bias across the domain. Virtually no spread (< 1 ppm) is found at SNP for both all hours and afternoon only, with very little changes between the two periods, consistent with its rural location and lack of enhanced CO_2 values from emissions sources.

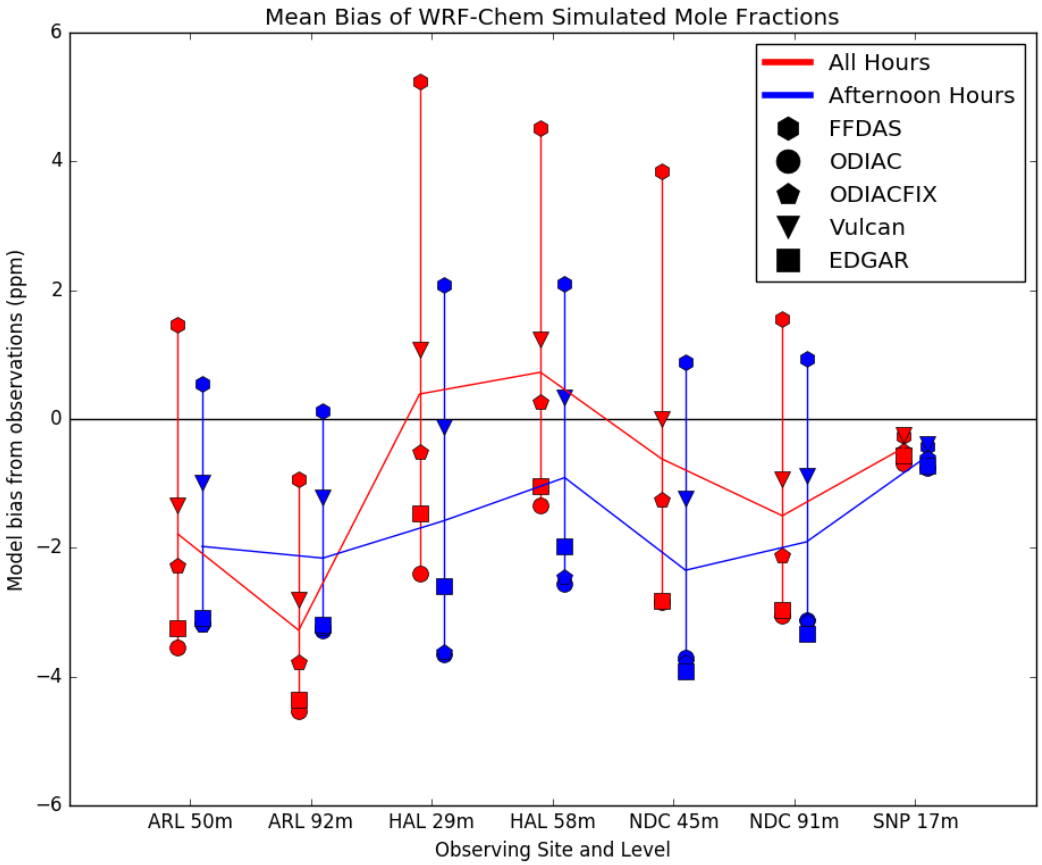

Figure 3.6. Mean bias of WRF-Chem simulated CO_2 mole fractions (ppm) compared to observations at each observing site and for both inlets where applicable for all five tracers (FFDAS: square, ODIAC: circle, ODIACFIX: pentagon, Vulcan: triangle, and EDGAR: square) during all hours (red) and during afternoon (12 pm to 4 pm local time; blue) hours only. Means of bias at each inlet are connected with additional lines: red for all hours and blue for afternoon.

3.3.2 Typical Meteorological Scenarios

The model-data comparisons presented in Section 3.3.1 are for simulated data over the entire month. While the mean biases of the modeled CO_2 mole fractions are relatively small, the variation between days, and even between individual hours of the day, can be significant, as shown in the residual plots in Fig. 3.5. Two sample cases (the dark gray shaded regions in Fig. 3.5) are presented to show a range of scenarios.

One of these cases is when the model error is large (greater than 10 % of the observed total value) and the proportion of variability (the ratio of the standard deviation of the differences over the period of interest and the standard deviation over the entire month) is greater than 100 %. The second scenario occurs when the model-observation difference is small (~<1 % of the total CO_2) and the proportion of variability relative to the whole month is under 40 %. For trace gases in the atmosphere, winds (both speed and direction) as well as the height of the PBL are the most important meteorological factors in estimating near-surface CO_2. In the subsequent subsections, two scenarios are presented that show examples of how different synoptic weather situations can affect the ability of WRF to predict these variables, and in turn, can impact the quality of the predicted CO_2 in a forward transport model.

3.3.2.1 Frontal Passages

A typical mid-latitude location such as the NEC-B/W experiences frontal passages every 3 to 7 days. When a front passes over an area, such as this urban domain, it causes sharp changes in wind speed and direction, temperature and moisture content, as well as other defining features of an air mass including CO_2 mole fractions (Parazoo et al., 2008). Figure 3.7 shows an example of the impact on CO_2 mole fractions when two fronts pass over the NEC-B/W as a mid-latitude cyclone moves east. The gradual slope of the warm front (~12 UTC February 3) can create a shallow PBL, allowing CO_2 to accumulate near the surface, whereas a cold front (~8 UTC February 4) is much steeper in its vertical structure. For the latter, the observed wind shifts are much more abrupt both in speed and direction. Both frontal passages

can also create surface convergence ahead of the front, allowing CO_2 to build up in the PBL immediately before the wind shift occurs. Figure 3.8 shows surface CO_2 concentrations predicted using the FFDAS emissions inventory and 10 m wind vectors during the model simulated frontal passage (5 UTC February 4), illustrating the spatial gradient of CO_2 during this period.

For both the warm and cold fronts, WRF simulates the frontal passage and associated wind shift 4 to 5 hours before the observed passage (Fig. 3.7). Generally, correlations are observed between the wind direction, the PBL height, and the simulated CO_2. The simulated warm front caps the PBL, allowing for predicted mole fractions to increase across the domain. The diurnal range is well resolved by WRF-Chem at NDC, but it is greatly overestimated at HAL and underestimated at ARL, likely related to the timing and position of the passing front. The spread between the various inventories at HAL is also much larger than at the other two sites. During the first frontal passage (12 UTC February 3), it is clear that the predicted CO_2 values are influenced by the shallow PBL depth, which is also observed at a number of periods during the month where the model and observations diverge. However, during the simulated cold front (~8 UTC February 4), the PBL heights do not change significantly, but the predicted CO_2 peaks and then drops rapidly as the wind shifts from southerly to northerly. This feature is seen in the simulated time series at all three sites with WRF-Chem results underestimating CO_2 mole fractions at the observing locations when the front actually passes through. This example illustrates that the meteorological error, in both the timing of the front and the PBL depth, dominates the error in the simulated CO_2 values rather than the underlying emissions,

as often the spread in the emissions inventories (shaded red area) is generally smaller than the difference between the model mean (red line) and the observations (black).

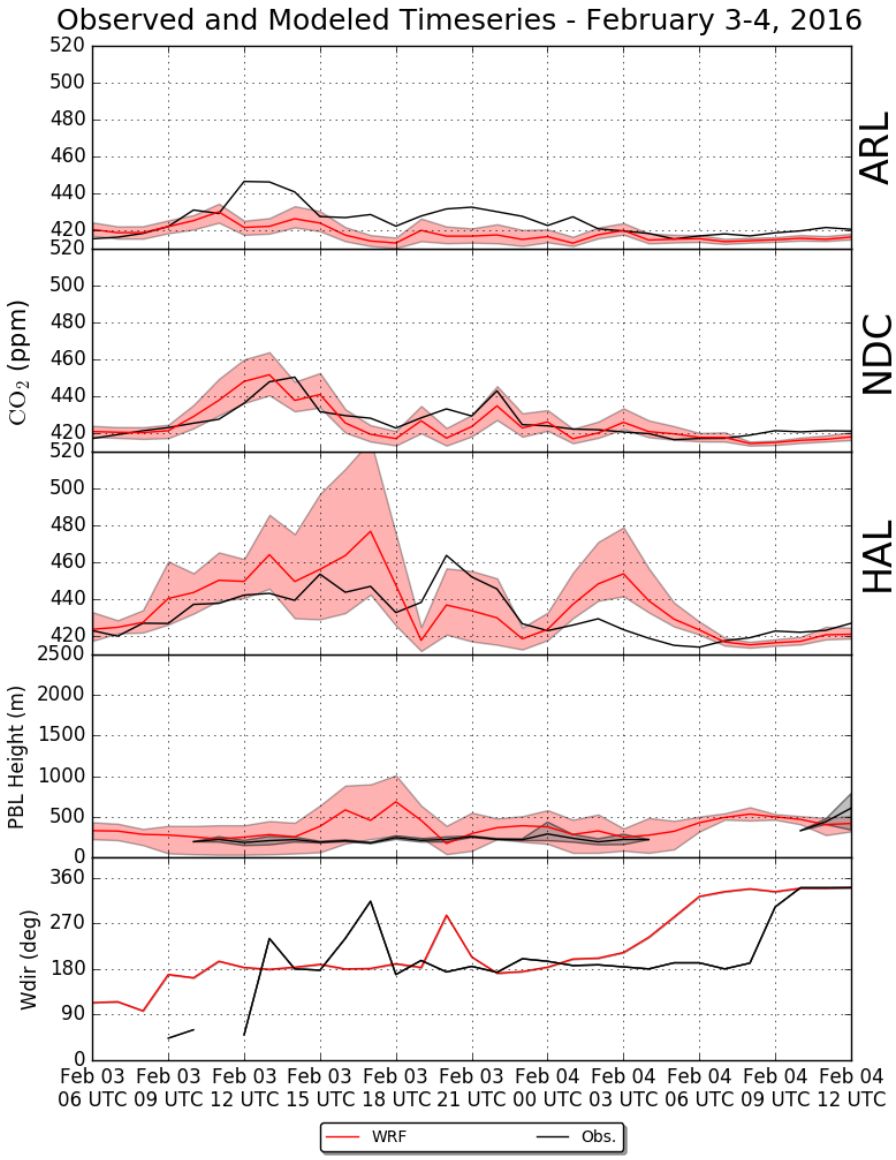

Figure 3.7. Observed (black) and modeled (red line average; shaded red spread of the five emissions inventories) hourly averaged CO_2 at all three urban sites for a typical frontal passage period (February 3-4, 2016). Bottom panels show observed (black line average of all observing sites; shaded spread of observations) and modeled (red line average of all observing sites; shaded spread of the modeled values at each observing site) hourly averaged PBL height, and 10 m observed (black) and modeled (red) wind direction at KDCA.

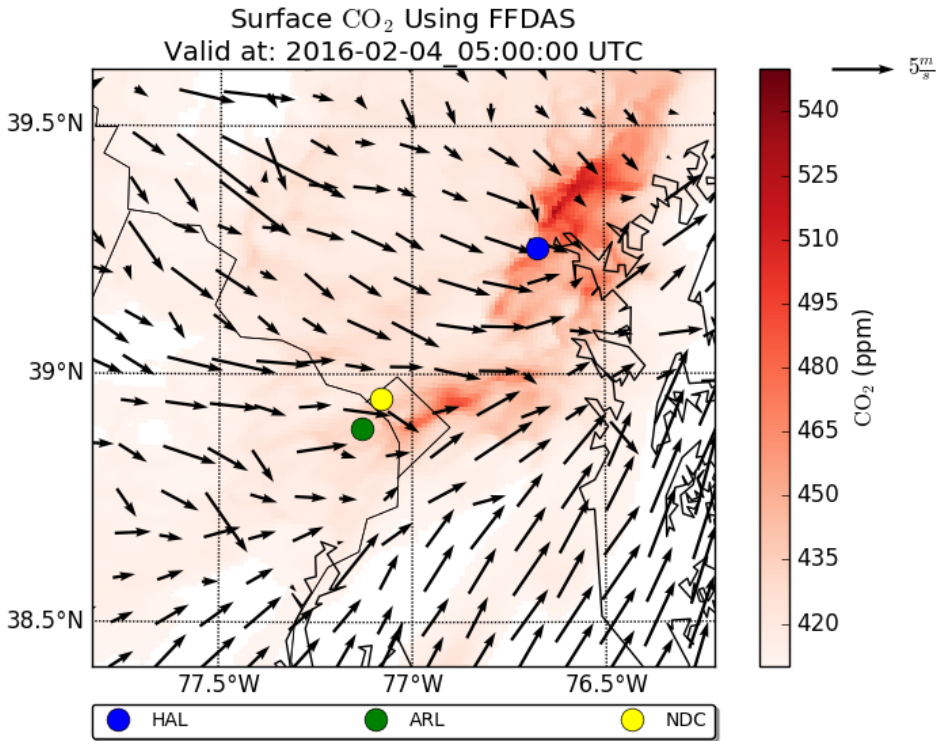

Figure 3.8. Simulated surface CO_2 concentrations using FFDAS emissions and 10 m wind vectors during a cold front passage at 5 UTC February 4, 2016 in domain d03. Locations of the three urban observing sites are shown (HAL in blue; ARL in green; NDC in yellow).

3.3.2.2 Persistent Winds

While the first scenario presents an example period where WRF-Chem significantly under or overpredicts CO_2 compared to the observations, there are times when the model simulated CO_2 is within 1 % of the observed value. One such example is from February 10 to February 11, 2016, when winds are steady and from a direction where the upwind CO_2 mole fractions are more representative of the global average. Figure 3.9 shows the modeled and observed CO_2 as well as wind direction and PBL height for this period. During these two days, the wind is persistently from the west or northwest, bringing in a steady stream of air into the urban area where the

mole fraction observations are strongly influenced by the incoming atmosphere, or regional background, values.

The average modeled CO_2 mole fractions at all three sites for this 48 hour subset are always within 5 ppm of the observed value. At ARL and HAL most hours are underpredicted, possibly due to the overprediction of the PBL height during this period as the modeled wind speed and direction agree well with the observations, but at NDC the average modeled value differs by 1 ppm to 2 ppm from the observations from 0 UTC to 18 UTC on February 11. Additionally, the predicted CO_2 mole fractions from all five inventories do not vary significantly from one another during this case, with a spread of only 2 ppm to 3 ppm on either side of the mean throughout the period. Despite this agreement, there are still some subtleties that can be observed in the data. All three sites have a local maximum in the observations at around 20 UTC on February 10, but the modeled time series have a local minimum there. At the same time, WRF overpredicts the PBL height compared to the observed height, likely causing dilution in the predicted CO_2 mole fractions. These features are present in all the inventory tracers, and often the observations are not within the spread of the five models, even during this period of relatively good model performance. In addition to the meteorological errors described above, biases in the CarbonTracker background or the biospheric tracer (both common to all five tracers) could also contribute to the overall error, but are likely limited to the magnitude of the bias observed at SNP.

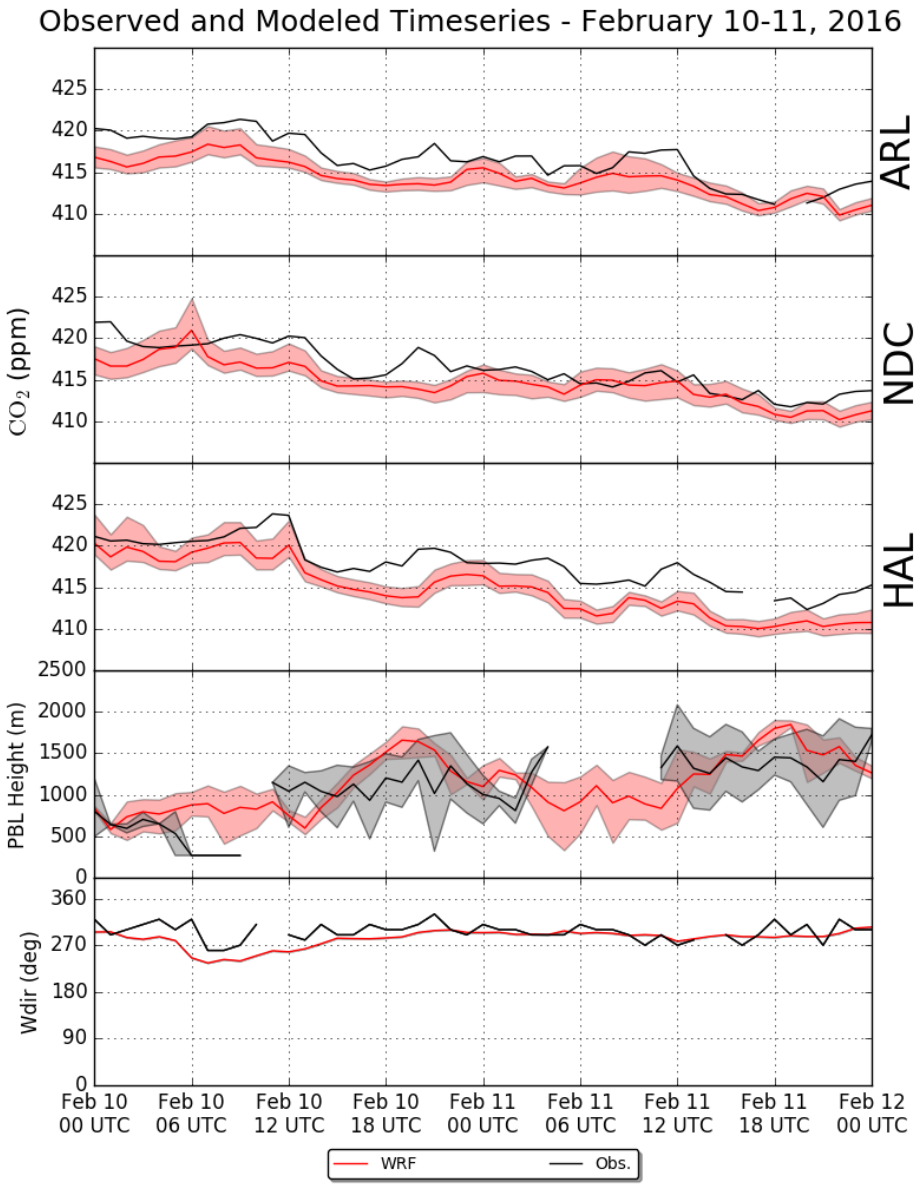

Figure 3.9. Observed (black) and modeled (red line average; shaded red spread of the five emissions inventories) hourly averaged CO_2 mole fractions at all three urban sites for a typical period with persistent winds from a rural area (February 10-11, 2016). Bottom panels show observed (black line average of all observing sites; shaded spread of observations) and modeled (red line average of all observing sites; shaded spread of the modeled values at each observing site) hourly averaged PBL height, and 10 m observed (black) and modeled (red) wind direction at KDCA. This shows that WRF-Chem is able to resolve both the wind direction and height of the PBL with reasonable skill, although deviations do occur.

3.4 Discussion

As discussed in Section 3.3.1, WRF-Chem tends to underestimate hourly-averaged values of near-surface CO_2 mole fractions when compared to observations averaged over the entire month. The daily maxima in simulated mole fractions from certain emissions inventories are often high relative to observations, particularly when they occur during the overnight hours when the modeled PBL depth is under predicted. However, the larger range of predicted values during these periods as well as the overall underprediction during the afternoon hours result in lower averaged values from the model than from observed values when including all sites, hours, and inventories as shown in Fig. 3.6. Previous work comparing simulated CO_2 to observed time series in an urban region also found an overall low bias, with predicted levels over certain hours/days exceeding observed levels (Feng et al., 2016). There are some synoptic situations, e.g., February 10 and February 11 (Section 3.3.2.2), where persistent winds allow for minimal errors in predicted CO_2 across the domain over an entire diurnal cycle, not just in the afternoon hours. During the afternoon of February 10, in fact, WRF overpredicts the PBL height, and thus underpredicts the near-surface CO_2. On this day, the modeled CO_2 may be more representative of reality during the overnight hours than it is in the afternoon. Conversely, other days, such as February 3 (Section 3.3.2.1), with a passing mid-latitude cyclone and its associated fronts, yield much different results. During this case, because of the predicted wind shift timing and the magnitude of the PBL height varying from observations, WRF tends to either overestimate or underestimate near-surface CO_2 depending on the hour.

In addition to the synoptic meteorology, variations in emissions inventories are also reflected in the predicted CO_2 mole fractions as shown in the mean biases described in Fig. 3.6. For the outermost domain (d01), the areal sum of the hourly averaged emissions is similar in magnitude (all within 10 % of the mean of the 4 inventories) (Fig. 3.3; Table 3.3). However, in the innermost domain (d03), the areal sum of FFDAS is over 36 % higher than the mean and the lowest inventory EDGAR is 20% below the mean of the inventories. Even though on the national scale each inventory is similar, there can be substantial differences between them due to the emission disaggregation methods (e.g., Hutchins et al. 2016; Oda et al. 2018) when considering mesoscale modeling of CO_2, a problem being studied further in other works (e.g., Fischer et al., 2017). The differences in totals and local sources could also be attributed to differing methodologies and datasets included in each emissions inventory, including the exact location of point sources and grid cell locations, among other things. These differences are generally reflected in the simulated CO_2 levels, with FFDAS being the highest averaged over the entire month, and EDGAR the lowest.

Traditionally, atmospheric inversions utilize data and meteorological model output from afternoon hours (12 pm to 4 pm local time) only (Kort et al., 2013; Breón et al., 2015; Lauvaux et al., 2016; Sargent et al., 2018). This is because the PBL is generally considered to be well-mixed during this time, and the model meteorology is thought to perform best compared with observations. By including prior emissions inventories, these inversions are used to estimate the total flux of carbon from an area of interest. However, in the forward modeling framework presented here with WRF-

Chem, using afternoon hours only may not be the best metric to determine whether the model meteorology is sufficient to accurately predict CO_2 mole fractions. Figure 3.10 shows the mean absolute error (the dark bars) as well as the mean standard deviation (the lightly colored bars) of the five predicted CO_2 mole fractions for each time series, both for all hours (red) and afternoon hours only (blue). When considering only the three urban sites, the mean absolute error for all hours is between 3.6 ppm and 7.0 ppm across the different observing sites and inlets whereas the mean standard deviations are between 1.9 ppm and 4.1 ppm. When including afternoon hours only, the mean absolute error of the model does decrease by an average of 1.42 ppm, and the standard deviations decrease by an average of 0.58 ppm. Despite the improvement in MAE when only including afternoon hours, the mean error of the model is still approximately twice the variation in the predicted CO_2 values from each emissions inventory. This result indicates that although limiting inversion analysis to afternoon hours may reduce overall meteorological model error it also limits analysis to time periods when local and regional emissions influence the observations the least (due to deeper PBLs and stronger mixing).

Additionally, the mean absolute error is roughly a factor of two larger than the mean standard deviation at the urban sites, with the ratio of the two ranging from 1.64 to 2.58 for all hours, and 1.57 to 1.89 for afternoon hours only, depending on the site and inlet level. This result suggests that on average, factors common to all five tracers (meteorological error, background error, or error in the biosphere tracer) contribute more to the overall model performance than the choice of anthropogenic emissions inventory. Given the low bias (~2 ppm) at SNP, the extent of our largest WRF

domain, and small relative contribution of the VEGAS tracer to the monthly mean values, we expect that the errors shown in Fig. 3.10 are dominated by meteorological conditions during winter. This conclusion is further supported by the two examples illustrated in Section 3.3.2. However, it is important to note that these meteorological conditions or errors can exacerbate the differences in the emissions inventories as well, like shown in Fig. 3.7.

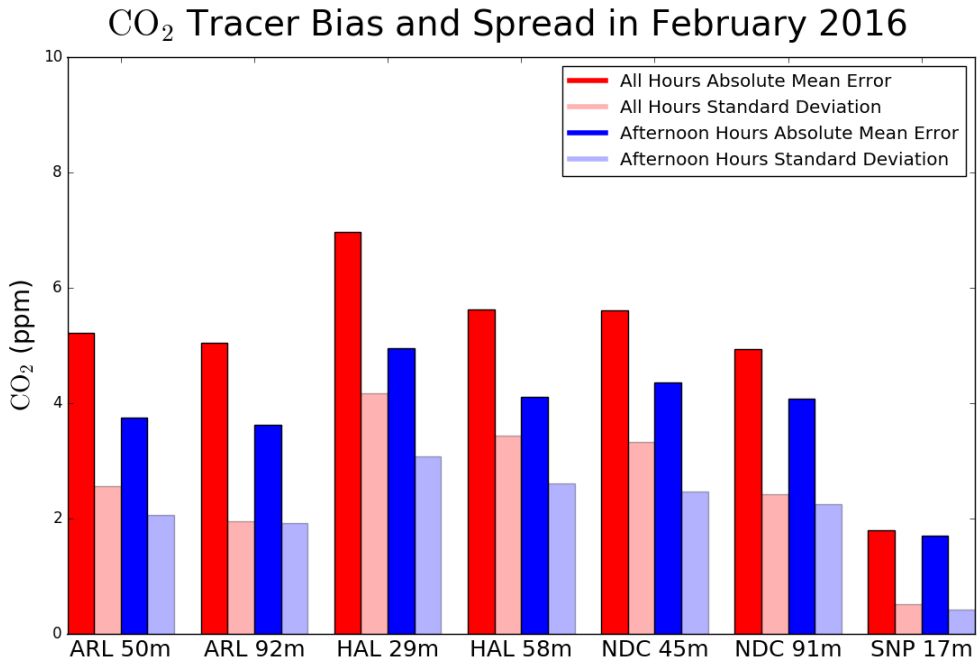

Figure 3.10. The mean absolute error (the dark bars) and the mean standard deviation (the lightly colored bars) of the five predicted CO_2 mole fractions for each observing site and inlet height (where applicable), both for all hours (red) and afternoon hours only (blue).

Rather than looking at the mean absolute error for the entire month and the mean standard deviation, it is also useful to consider the mean absolute error of the five tracers compared to the observed mole fraction at each hour and the associated standard deviation to get an idea of how the error and spread of the modeled CO_2 are

related. These two values are reasonably related with a correlation coefficient of approximately 0.3 across the three urban sites over the month. At times, generally when the modeled CO_2 is at its highest, which is when transport model errors tend to be largest such as during the overnight hours or frontal passages, approximately a factor of two difference between the highest and lowest modeled enhancements can be found. But for many other cases, and on average as described above, the differences between the various emissions inventory tracers are smaller than the absolute error relative to observations. During periods of low modeled CO_2 error (i.e. small differences from the observations), the variation among the different emissions models is small but still discernable (as shown in Fig. 3.9 for example). Regardless, the error in modeled CO_2 compared to observations for any given hour or day appears to be influenced more by the meteorology than the differences among the various emissions inventories. This result may pose a challenge to an atmospheric inversion that adjusts fluxes to match observed concentrations when variability in the modeled CO_2 mole fractions is dominated by meteorological error. It also suggests the need for methods that identify time periods when meteorological conditions are best represented by simulations, rather than only including afternoon hours, and when it may be anticipated that inversions will perform with higher fidelity to actual emissions conditions. This is particularly important for urban applications where the impact of synoptic variability may strongly and adversely impact inversion analyses.

3.5 Conclusions

An evaluation of WRF-Chem simulated CO_2 mole fractions using multiple anthropogenic CO_2 emissions inventories at four CO_2 observing sites in the Baltimore, MD and Washington, DC metropolitan areas was presented above. For all emissions inventories the modeled CO_2 is within 5 ppm of observations when averaged over all observing sites for the month of February in 2016. However, for any given hour, at any particular site, the differences between the ensemble of simulated CO_2 values and the observed CO_2 can vary from near zero to as high as 100 ppm (Fig. 3.4). The differences between the simulated time series for the different emissions inventories vary significantly in time, but tend to be proportional to the magnitude of the enhancement over the background CO_2 value. When averaged over the entire month all simulated CO_2 mole fractions are within 8 ppm of each other (Fig. 3.6) representing a range of approximately 2% of the total mole fraction.

This analysis suggests that the predicted mole fraction error relative to observations is dominated by model meteorology and not the underlying emissions inventory in winter months when looking at individual observing sites. Not only do certain synoptic setups allow for minimum absolute errors in the predicted values, but the timing and location of frontal passages can significantly impact the model performance at predicting CO_2 mole fractions. We also find that the errors associated with atmospheric transport are not restricted to certain times of day. This suggests that filtering data based on model performance rather than time of day (such as using only mid-afternoon observations) for atmospheric inversions might yield better overall results and thus further methods, such as machine learning algorithms, are

needed to better identify time periods where the simulated transport performs well. To improve the simulated CO_2 mole fractions error relative to observations, the prediction of key meteorological variables such as wind speed and direction and the height of the PBL must be improved, either through more advanced physics schemes or through data assimilation techniques. As such, minimizing errors associated with atmospheric transport and dispersion generally will improve the performance of estimated fossil fuel CO_2 emissions more than improving emission priors.

Chapter 4 Sensitivity of Urban Carbon Flux Estimations to Observation Networks and Ensemble Data Assimilation System Configuration in Observing System Simulation Experiments

Chapter based on manuscript in prep for submission to Atmospheric Chemistry and Physics, Fall 2018.

Cory R. Martin, Ning Zeng, Russell R. Dickerson, Eugenia Kalnay, Anna Karion, Tomohiro Oda, Yun Liu, and James Whetstone

4.1 Introduction

As various geopolitical entities embark upon GHG mitigation efforts, accurate quantification of GHG emissions can significantly inform their effective management, in addition to providing quantitative substantiation of progress toward emission reduction goals. As mentioned in Chapter 1, observations of CO_2 mole fractions are used in conjunction with atmospheric transport models to estimate the total enhancement of greenhouse gases from the urban area and with statistical techniques to infer the underlying emissions. There are a number of examples of various statistical methods used for this flux estimation as well as different locations and spatial and temporal scales (e.g. Bousquet et al., 1999; Michalak et al., 2004; Gourdji et al., 2008; Mueller et al., 2008; McKain et al., 2012; Lauvaux et al., 2016*)*, including atmospheric data assimilation techniques such as variational methods and Kalman filters (Gerbig et al., 2003; Peters et al., 2005; Baker et al., 2006; Kang et al., 2011, 2012; Chatterjee et al., 2012; Chatterjee and Michalak, 2013; Liu et al., 2016).

Chapter 2 focused on the use of small, commercially available, low-cost sensors that measure CO_2 and indicated that after environmental correction, these

sensors demonstrate their promise for ambient air monitoring. Previous work has shown that an increase in observational density can improve quantification of fossil fuel emissions through statistical inversions (Turner et al., 2016; Lopez-Coto et al., 2017; Wu et al., 2018). However, because of the cost of the observing sites tend to be in excess of $100,000 each, most urban GHG monitoring networks feature fewer than a dozen or so locations, with some exceptions (Shusterman et al., 2016). By utilizing new technology such as these lower-cost CO_2 observing platforms, there is the potential to decrease the uncertainty in the total emissions of an urban area.

In this chapter, a series of observing system simulation experiments (OSSEs) are conducted using an ensemble Kalman filter data assimilation system along with a mesoscale meteorology model coupled with passive tracer transport to evaluate the sensitivity of the estimation of fossil fuel CO_2 emissions from the NEC-B/W to the quantity and quality of CO_2 in situ observations used as well as different ensemble configurations. Section 4.2 describes the transport model and the Kalman filter used. Section 4.3 details the experiment setup, including what is used as the assumed truth as well as the generation of synthetic observations. Section 4.4 describes results from the different experiments featuring different observation networks, ensemble inflation techniques, and observation localization radii. Section 4.5 features a discussion on the results of Sect. 4.4. Finally, conclusions are presented in Section 4.6.

4.2 Methods

For these perfect model OSSEs, a mesoscale weather prediction model including atmospheric transport of trace gases is modified to include 50 distinct three-

dimensional (3D) tracers using different anthropogenic fluxes to provide an ensemble of solutions with only one realization of meteorology. This allows for no assumed error in meteorological transport or dispersion, while at the same time generates an ensemble of CO_2 with relatively minimal additional computation cost compared to a single tracer model simulation. For each 3D CO_2 tracer, there is also a corresponding surface CO_2 flux variable used as a constant source or sink (depending on sign) in time between analysis cycles of CO_2 at the model surface. The ensemble data assimilation code has also been modified to interpret each 3D tracer in the model output as well as its respective surface flux variable as a discrete ensemble member for the analysis cycle.

4.2.1 Forecast Model

The Weather Research and Forecasting model coupled with chemistry (WRF-Chem) is used as the atmospheric transport and meteorological prediction model. It is a non-hydrostatic, compressible model that provides passive tracer transport online with mesoscale meteorology forecasting capabilities (Grell et al., 2005; Skamarock et al., 2008; Beck et al., 2011). A nested grid is defined with an outer domain featuring a horizontal resolution of 9 km and encompassing roughly the northeastern quadrant of the United States. The inner domain has a horizontal resolution of 3 km and is approximately 350 km in each direction centered over the Baltimore, MD-Washington, DC metropolitan areas. For both domains, 50 vertical levels are used and are chosen by WRF-Chem at initialization. Initial and boundary conditions of the meteorological variables are from the NOAA NCEP North American Regional

Reanalysis (NARR), a product with a horizontal resolution of 32 km, 30 vertical layers, and 3-hourly output (Mesinger et al., 2006). For the CO_2 tracer, initial and boundary conditions are provided by NOAA ESRL's CarbonTracker Near Real-Time (CarbonTracker-NRT; version 2016) gridded product (Peters et al., 2007; https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT-NRT/index.php). This is a 3D mole fraction product with 3-hourly output and a horizontal resolution of 1º over North America. The rest of the model configuration and domain information are identical to that described in Chapter 3 except in this case the innermost domain with 1 km horizontal resolution is not used, in order to lower the computational cost of each OSSE.

4.2.2 Data Assimilation

The Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007) is a type of ensemble Kalman filter where the forecast error covariance statistics are computed from the difference of each member forecast for any variable, **x**, ($\mathbf{x}^{f(i)}$ where *i* is the ensemble member number) from the ensemble mean forecast ($\bar{\mathbf{x}}^{f}$). The information from the forecast ensemble combined with observations ($\mathbf{y}^{o}$) results in a new analysis mean ($\bar{\mathbf{x}}^{a}$) and a new analysis for each ensemble member ($\mathbf{x}^{a(i)}$ where again *i* is the ensemble member number). In WRF-Chem, an online observation operator (*H*) has been implemented to interpolate the model state from every ensemble member to each observation at its specific time and location $\mathbf{y}^{f(i)} = H\mathbf{x}^{f(i)}$ as the ensemble forecast integrates, resulting in both $\mathbf{x}^{f(i)}$ and $\mathbf{y}^{f(i)}$ being computed

at the same time. For the simplest case of in situ observations of the variable being assimilated, as is the case of CO_2 mole fraction point observations, H is a three-dimensional interpolation of the model variable to the observation point in space.

LETKF generates the analysis at every grid point independently of one another, allowing for efficient parallelization. This is achieved by only including observations within a certain localization radius from the computed grid point. The analysis mean at each grid point in the model is shown in Eq. (4.1):

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}^f \bar{\mathbf{w}}^a \tag{4.1}$$

where $\mathbf{X}^f$ is the matrix of the difference of each ensemble member forecast from the forecast mean and $\bar{\mathbf{w}}^a$ are the weights applied to each ensemble member for the analysis, with $\mathbf{R}$ denoting the observation error covariance matrix, $\tilde{\mathbf{P}}^{\mathbf{a}}$ is the analysis error covariance, K is the number of ensemble members, $\mathbf{I}$ is the identity matrix, and $\mathbf{Y}^{\mathrm{f}}$ is the perturbations of the ensemble forecast from the ensemble mean in observation space:

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^{\mathrm{a}} (\mathbf{Y}^f)^T \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^f) \tag{4.2}$$

$$\tilde{\mathbf{P}}^{\mathrm{a}} = [(\mathbf{Y}^f)^T \mathbf{R}^{-1} (\mathbf{Y}^f) + (K-1)\mathbf{I}]^{-1} \tag{4.3}$$

$$\mathbf{Y}^f = \mathbf{y}^{f(i)} - \bar{\mathbf{y}}^f \tag{4.4}$$

Previous work has shown that the LETKF can be successfully applied to estimate surface carbon fluxes with atmospheric CO_2 mole fractions observations

used and both CO_2 fluxes and mole fractions used as control variables (Kang et al., 2011, 2012; Liu et al., 2016). However, these examples are applied to a global domain with a coarse resolution model, and not a mesoscale model with sufficiently high resolution (sub 10 km) such as WRF. For the experiments evaluated in this study, the data assimilation system is cycled every three hours, generating a new analysis of the three-dimensional CO_2 field and the surface CO_2 flux each time. While observations of CO_2 flux are not used, by assuming in the LETKF system that surface flux and the CO_2 mole fraction are correlated (variable localization) we can solve for both control variables using only mole fraction observations. Additionally, while the three-dimensional CO_2 field is optimized for both WRF domains, only the input for d02 is adjusted by LETKF for the surface CO_2 flux.

4.3 OSSE Design

An observing system simulation experiment, or OSSE, is often used to evaluate how a new set of observations or a new technique can improve a model analysis, and possibly the resulting forecast. Normally, data assimilation experiments are compared to another analysis, and use that as the assumed correct or reference dataset, but there is still some error associated with this analysis. In an OSSE, the true state of the system can be known as it is generated by a previous model simulation. Observations are created from this previous model simulation, which is often called the “nature run”. The following subsections describe how the nature run is performed, the observations generated from the nature run, and the experiments that will be evaluated in later sections.

4.3.1 Nature Run

First, WRF-Chem is initialized with NARR meteorology at 00 UTC February 01 2016 and uses CarbonTracker-NRT as the initial conditions of CO_2 as described in Sect. 4.2.1. Then the model is integrated for the first ten days of February 2016 with the CarbonTracker-NRT boundary conditions as well as a prior emissions inventory to generate a 4D field of CO_2 across both model domains for that period. This length of time was chosen as it allows for multiple diurnal cycles and 1-2 synoptic cycles while at the same time allowing for relatively low computation cost and time for multiple experiments. Synthetic observations are generated from this 4D CO_2 dataset (henceforth referred to as "nature") and used for the subsequent ensemble data assimilation experiments. Details on the assumed true fluxes, synthetic observations, and the ensemble generation are provided in the following subsections.

4.3.1.1 Fluxes Used in Nature Run

The Open-source Data Inventory for Anthropogenic CO_2 (ODIAC; Oda and Maksyutov, 2011; Oda and Maksyutov, 2015; Oda et al., 2018), is used as the emissions inventory for fossil fuel CO_2 for the nature simulation. ODIAC was chosen because its horizontal resolution of ~1 km is higher relative to other available emissions inventories. Using the total emissions estimated by the Carbon Dioxide Information and Analysis Center (CDIAC) at the US Department of Energy's Oak Ridge National Laboratory, the location of point sources and satellite-derived nightlights are then used to distribute the emissions onto the ODIAC grid. Monthly total fluxes are provided by ODIAC for each month projected using statistical data

from the energy company BP for most recently, the year 2015. The emissions inventory is linearly interpolated from its native grid to the two domains for use as input in the WRF-Chem model, and the fluxes used for domain d02 are shown in Fig. 4.1. The ODIAC provided fluxes are hourly average fluxes for the entire month with no diurnal or day of week variability, and no temporal scaling was added as to make the true fluxes constant throughout each experiment, for simplicity in comparing to estimated values. As ODIAC was generated for 2015 but the meteorology is for 2016, ratios computed from the U.S. Energy Information Administration's (EIA) Monthly Energy Review's total anthropogenic CO_2 emissions (https://www.eia.gov/totalenergy/data/monthly/) are used to scale each of the emissions products as shown in Equation 4.5. The ratio of the totals provided by the EIA for February 2016 compared to February 2015 is applied as a constant scaling factor across the entire inventory.

$$ODIAC_{2016} = ODIAC_{Feb\ 2015} \times \frac{EIA\ Total_{Feb\ 2016}}{EIA\ Total_{Feb\ 2015}} \tag{4.5}$$

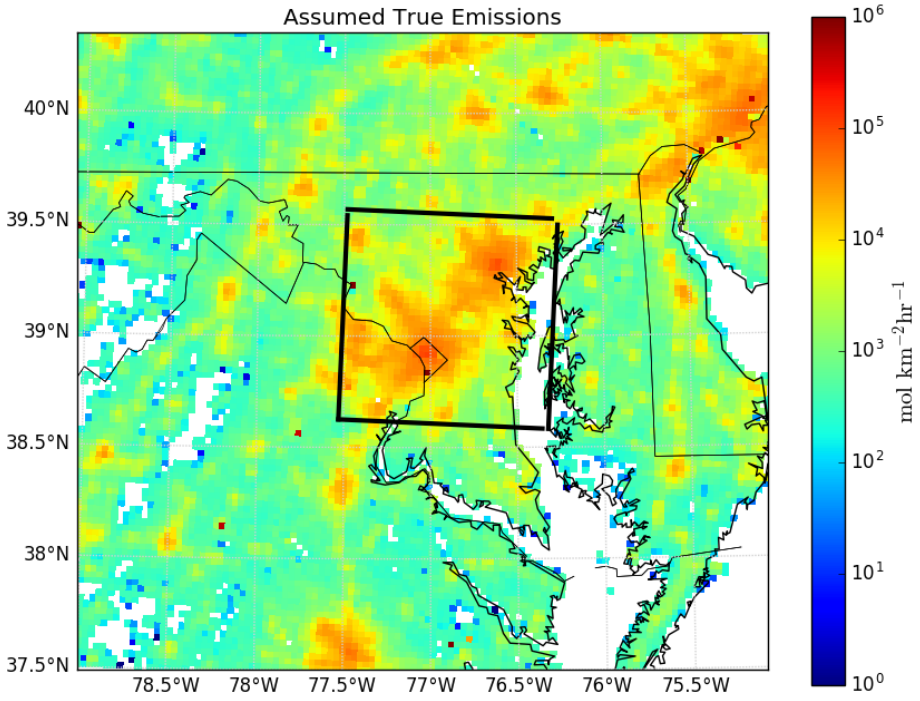

Figure 4.1. Map showing the assumed true anthropogenic emissions of CO_2 used in the nature run. The black lines draw a box 105 km x 105 km square centered over Baltimore, MD and Washington, DC to represent an "urban" region for flux estimation comparisons. Note that the color bar is a logarithmic scale from 1 to 1,000,000 mol km^{-2} hr^{-1}.

In addition to the fossil fuel emissions, a dynamical vegetation model has been coupled to WRF-Chem to provide the contribution of biogenic fluxes to the synthetic observations. The VEgetation-Global-Atmosphere-Soil (VEGAS) model (Zeng et al., 2005) is coupled offline with WRF-Chem to provide hourly biospheric CO_2 flux. To generate the land to atmosphere carbon flux, VEGAS uses the WRF-Chem meteorological output variables (2 m temperature, 2 m specific humidity, hourly precipitation, 10 m winds, skin temperature, and total net radiation) as well as the WRF domain topography, emissivity, and albedo. As this simulation is for February 2016, the biogenic component is relatively small compared to the fossil fuel emissions but is included anyway in the model to most accurately represent the CO_2

mole fractions observed in the mid-latitude winter. Because the model meteorology is fixed for all ensemble members and considered perfect relative to the nature simulation, the biospheric fluxes are not adjusted by the LETKF, but are generated again and included in the simulated CO_2 mole fractions for the assimilation experiments.

4.3.2 Creating Synthetic Observations from Nature Run

To understand the effect the quality of observations has on the CO_2 flux estimation, two types of in situ observations of CO_2 mole fractions are generated from the nature run. Both types of observations are linearly interpolated to a height above ground level (AGL) and are an hourly average of the previous hour's instantaneous model output at every ten minutes. One of the types of observations is a high-accuracy instrument with an inlet on a tower, generally with sufficient height AGL, which will be called Obs_HA referring to high-accuracy. The other is a medium-precision assumed to be low-cost sensor generally located at a lower height (~10 m AGL) comparable to that of typical buildings in an urban/suburban area which will be referred to as Obs_LC (for low-cost). The error assumed with each observation is calculated from Equation 4.6:

$$(4.6)\ Error_{Obs} = \sqrt{(\frac{\sigma(CO_2)}{\sqrt{n}})^2 + Error_m{}^2}$$

where $Error_{Obs}$ is the total error associated with each hourly observation, $\sigma(CO_2)$ is the standard deviation of the instantaneous model output contained in the hourly

average, n is the number of model output times in the average (in this case 6, every 10 minutes within an hour), and $Error_m$ is the measurement error assumed of the instrument. For the high-accuracy locations (Obs_HA), $Error_m$ is assumed to be 0.2 ppm, and for the medium-accuracy sensor locations (Obs_LC), $Error_m$ is 2 ppm. Four different observing network configurations are tested in subsequent experiments featuring both of these types of aforementioned observations: 1) NEC-B/W towers, 2) low-cost sensors, 3) hybrid, and 4) ideal. Each is described in detail below. Table 4.1 summarizes these four networks, and maps of their spatial distribution are available in Fig. 4.2.

4.3.2.1 NEC-B/W Network

For the first network, named NEC-B/W, the locations of either existing or planned sites from the NIST NEC-B/W GHG observing network are used, with each assumed to be a high-accuracy observation point (Obs_HA as defined above). Each of these observing sites has a different height above ground level, and the WRF-Chem observation operator (H) uses the actual inlet height for the towers in the NEC-B/W. This network features 20 locations, with some in urban locations and the rest in the surrounding rural areas approximately halfway between the city centers and the edge of domain d02.

4.3.2.2 Low-Cost Network

Because the medium-accuracy sensor is defined earlier (Obs_LC) to have error an order of magnitude larger than the tower sites, we choose to test a network using 200 such sensors rather than 20 tower sites. These are distributed in the domain d02 by randomly selecting the center of 200 urban pixels and at a height of 10 m, simulating a distribution of sensors that would most likely be on rooftops or light posts. This network is referred to throughout the remainder of this Chapter as the "Low-Cost" network.

4.3.2.3 Hybrid Network

The third type of observation network evaluated combines all of the observation sites from the first two networks, i.e. both towers and sensors or what will be called in this Chapter, the hybrid network. Specifically, the locations and types of observations are the same as the first two networks, meaning 200 Obs_LC observations in urban areas and 20 Obs_HA observations in the actual NEC-B/W network locations.

4.3.2.4 Ideal Network

Finally, to test the viability of these data assimilation methods using as many observations as possible, an ideal network is generated featuring an Obs_HA observation at a height of 5m at the center of each grid point in domain d02, i.e. approximately 3 km apart. This results in 13,320 high-accuracy observations in the

domain. While this is impractical from a cost or logistical standpoint, it is purely an experiment to see how the system can perform with the highest level of constraint of in situ observations.

Table 4.1. Summary of the four different observing network configurations used in this study.

Network Name	NEC-B/W	Low-Cost	Hybrid	Ideal
# of observing sites	20	200	220	13320
# of Obs_HA	20	0	20	13320
# of Obs_LC	0	200	200	0
Observation height (above ground level; m)	17-152	10	10-152	5

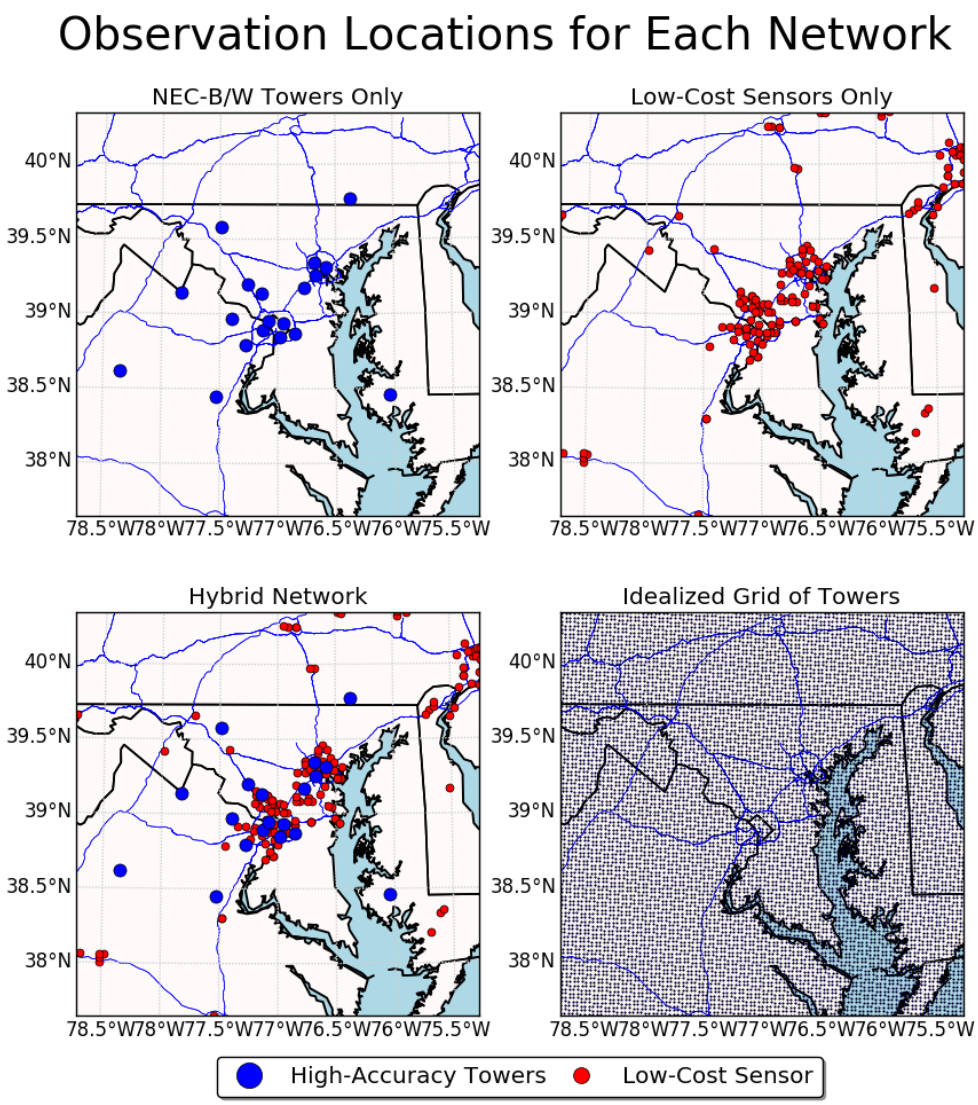

Figure 4.2. Map showing the distribution of observations for each of the four different in situ network configurations used in this study.

4.3.3 Ensemble Data Assimilation Configurations to Test

4.3.3.1 Ensemble Initialization

For all of the subsequent experiments, the 50 distinct CO_2 tracers comprising the ensemble are each initialized with a random 3D snapshot in time from the nature simulation output. This is done to provide a reasonable mean and spread of CO_2

encompassing the variations occurring on both the diurnal and synoptic time scales. For each tracer ensemble member, their respective surface flux arrays are initialized with a constant value of 0 mol km^{-2} hr^{-1} across the entirety of domain d02 so that no prior information or initial spread of the surface flux is given to the model. The assumed true fluxes from ODIAC are provided for domain d01 and are not adjusted by the data assimilation system.

4.3.3.2 Ensemble Inflation

After each assimilation cycle, the new ensemble analysis will tend to have an underestimated uncertainty because of limitations in the number of ensemble members as well as model error. If this continues, this can lead to filter divergence and sub-optimal results. To account for this, in ensemble data assimilation the individual members are inflated using various techniques to introduce additional spread, and thus uncertainty, to the ensemble. We chose to use additive inflation for both the CO_2 and flux fields, as it prevents the ensemble from trending toward the dominant direction of ensemble growth (Whitaker et al., 2008; Kalnay et al., 2007) and allows for a relatively easy method for adjusting the value for each variable and grid point.

$$Analysis_{final}(m) = Analysis_{LETKF}(m) + (Nature(t,m) - \overline{Nature}) * C \quad (4.7)$$

Equation 4.7 describes how additive inflation is performed on each of the 50 CO_2 tracer ensemble members. For any member, *m*, the 3D field after LETKF assimilates observations ($Analysis_{LETKF}$) has a perturbation added to it to inflate the

total spread of the ensemble. To get this perturbation, a random time is chosen from the Nature run (t) selecting a 3D CO_2 field for each member (m). The mean of these samples is computed, and then each member's 3D field has the mean subtracted from it to compute the perturbations about the mean. Finally, before they are added to the analysis fields for the final analysis for each member, the perturbations are multiplied by a coefficient (C), to scale the final inflation added at each analysis cycle. For the majority of this Chapter, the coefficient C is 0.1. Increasing or decreasing this scaling coefficient would increase or decrease the percentage of the perturbations added to each member and thus increase or decrease the ensemble spread.

(4.8) $$Analysis_{final}(m) = Analysis_{LETKF}(m) + (B * C)$$

Equation 4.8 shows how the additive inflation is applied to the CO_2 mole fraction ensemble members but the general form can be used for the 2D flux ensemble members as well. However, this equation can be put in a more general form where the additive inflation added to the initial analysis for each member is the product of the scaling coefficient, C (again assumed to be 0.1 unless otherwise stated), and an initial perturbation, B. In order to test the sensitivity of the ensemble composition to the final flux estimation, three different techniques for generating the initial perturbation (B) for each flux ensemble member are evaluated. Results comparing these different techniques will be presented in Sect. 4.4.1 and each will be described in the following subsections.

4.3.3.2.1 Infl_Constant

(4.9) $B_{Infl\_Constant} = R(m) - \bar{R}$

The simplest, which will be referred to as Infl_Constant, is a constant value applied to each pixel in the domain, meaning there is no variation in space. For each ensemble member (m) at each analysis time, a random value R between 1 and 5000 mol km^{-2} hr^{-1} is chosen. Then the mean of the 50 members is calculated, and the perturbation for each member about this mean equals B in Equation 4.8. This inflation technique was chosen as it introduces no prior spatial information into the estimated flux analysis.

4.3.3.2.1 Infl_Truth

(4.10) $A(m, x, y) = ODIAC(x, y) * a$

(4.11) $B_{Infl\_Truth} = A(m, x, y) - \overline{A(x, y)}$

Unlike Infl_Constant, which introduces no spatial information, Infl_Truth uses the assumed fluxes from the nature run to generate the additive inflation values. For each ensemble member at each analysis cycle, a random coefficient, a, between 0.001 and 0.5 is selected and the ODIAC emissions are then multiplied by this coefficient, creating an intermediate scaled 2D flux field, A (Eq. 4.10). Again, to get the value added (or subtracted) to the ensemble member after the analysis, the mean of these scaled fluxes is computed, and the difference from the mean for each

ensemble member is what is added to the analyzed fluxes for that member (Eq. 4.11). This inflation method allows for not only a realistic spatial pattern, but can also provide information for the relative magnitudes of fluxes at adjacent grid points.

4.3.3.2.3 Infl_PtSrc

$$A(m,x,y) = \begin{cases} ODIAC(x,y) > 100000 = ODIAC(x,y) * a \\ ODIAC(x,y) \leq 100000 = R \end{cases} \quad (4.12)$$

$$B_{Infl\_PtSrc} = A(m,x,y) - \overline{A(x,y)} \quad (4.13)$$

A compromise between the two techniques, Infl_PtSrc is also evaluated. At this high spatial resolution, particularly in populated areas, anthropogenic fluxes of surface CO_2 can vary by orders of magnitude from one pixel to the next, due to power plants or other point sources. This third inflation method is an attempt to constrain these point sources to be large values (both in uncertainty and mean flux), while keeping the rest of the domain free from any prior information. This is a relatively reasonable assumption as one may have information on the location of point sources (and even possibly a reasonable accurate estimation of their emissions) but not know where smaller sources are located. For most of the domain this method is the same as Infl_Constant, a random value, *R*, between 1 and 5000 mol km^{-2} hr^{-1} is chosen randomly. However, for pixels in the true fluxes that are greater than 100,000 mol km^{-2} hr^{-1}, the value is replaced by the ODIAC pixel's value multiplied by a random coefficient, *a*, between 0.001 and 0.5 (Eq. 4.12). As with the other two methods, the

mean is then computed and the perturbation B is the difference of A for each ensemble member from the mean of all of the A arrays computed (Eq. 4.13).

4.3.3.3 Localization Radius

Section 4.2.2 described the LETKF data assimilation framework and associated equations and mentioned that the localization of LETKF is applied to each model analysis grid point. To remove spurious correlations at long distances, localization is applied to optimize results in the data assimilation analysis. In this system, one user configurable option is a localization radius, which is based on a Gaussian distribution, where the radius prescribed is equal to the value of 1-sigma in distance. This distance tells the LETKF system for any given grid point in the model which observations to assimilate for the analysis and their associated weights of influence. So, generally, the larger the localization radius, the more observations that will be included. Section 4.4.2 will investigate how varying this localization radius affects the results of the assimilation system.

4.4 Results

In the following subsections, a number of experiments are conducted to evaluate the ability for an ensemble data assimilation system featuring WRF-Chem to accurately estimate urban surface CO_2 fluxes. First, the three additive inflation methods described in Sect. 4.3.3 are compared for each of the observation networks. Next, the sensitivity to the observation localization radius, meaning the distance that an observation can influence the optimization of a single analysis grid point, is evaluated for different radii. Finally, we examine how the scaling coefficient applied

to the inflation perturbations, as well as the location and number of observations can affect the results.

4.4.1 Sensitivity to Additive Inflation Methods

In an ensemble square root filter such as LETKF, the forecast error covariance statistics are computed from the spread of the different ensemble members. Thus, the ability to adjust the surface flux for each grid point is related to the uncertainty of that pixel in each forecast at every analysis cycle. To evaluate how the different additive inflation methods described in Sect. 4.3.3.2 affect both the spatial distribution and magnitude of estimated fluxes, the three different methods are implemented for all four observing networks using the same observation localization radius (100km). More results with varying the localization radius will be provided in Sect. 4.4.2.

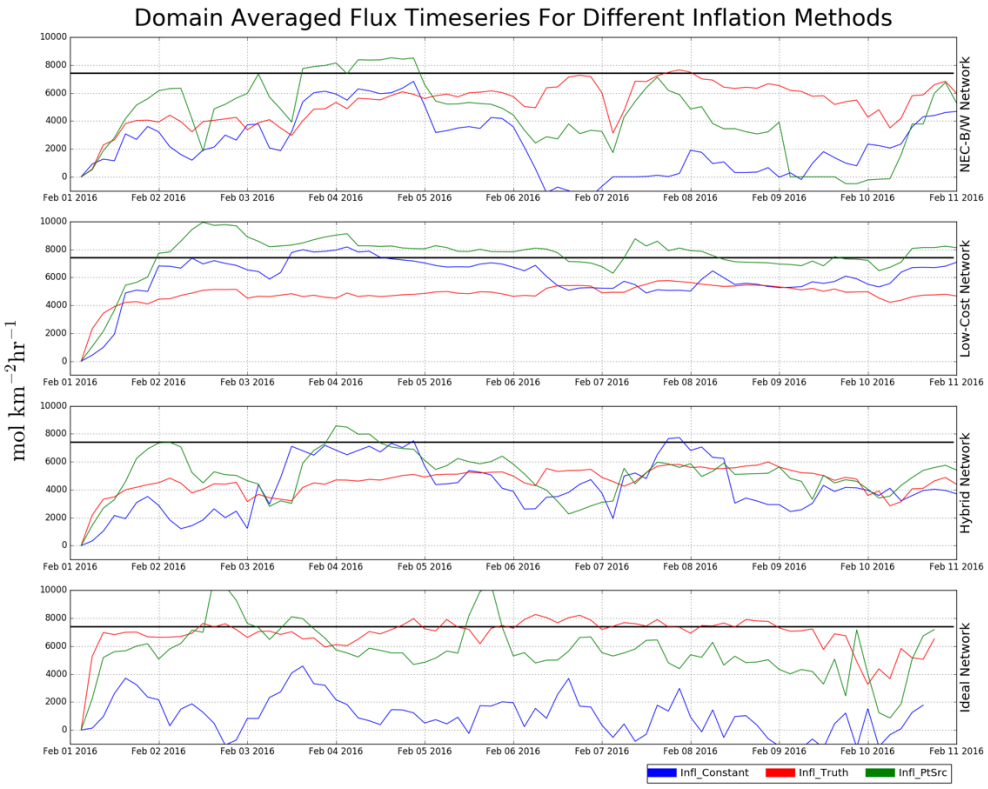

Figure 4.3. Time series of domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} for each of the four observing networks (Top: NEC-B/W; Top-Middle: Low-Cost; Bottom-Middle: Hybrid; Bottom: Idealized) and for each of the three additive inflation methods (Blue: Infl_Constant; Red: Infl_Truth; Green: Infl_PtSrc) and the assumed true average flux in black.

Figure 4.3 shows a time series of the average surface CO_2 flux for the nature run (black) as well as the three different additive inflation methods for each of the four observing networks for domain d02. For the domain averaged estimated surface flux, there is significant variability in each of the time series depending on which observing network is used and which inflation method is applied. This variability is not only apparent when changing the observation network or inflation technique, but depending on the aforementioned choices, can be significant in time as well. Some of these variations are correlated across different experiments and even observing networks, such as the overall decrease in most of the experiments over the last 1-2

days or the sudden drops observed in all three of the NEC-B/W experiments on February 2-3 and February 6-7.

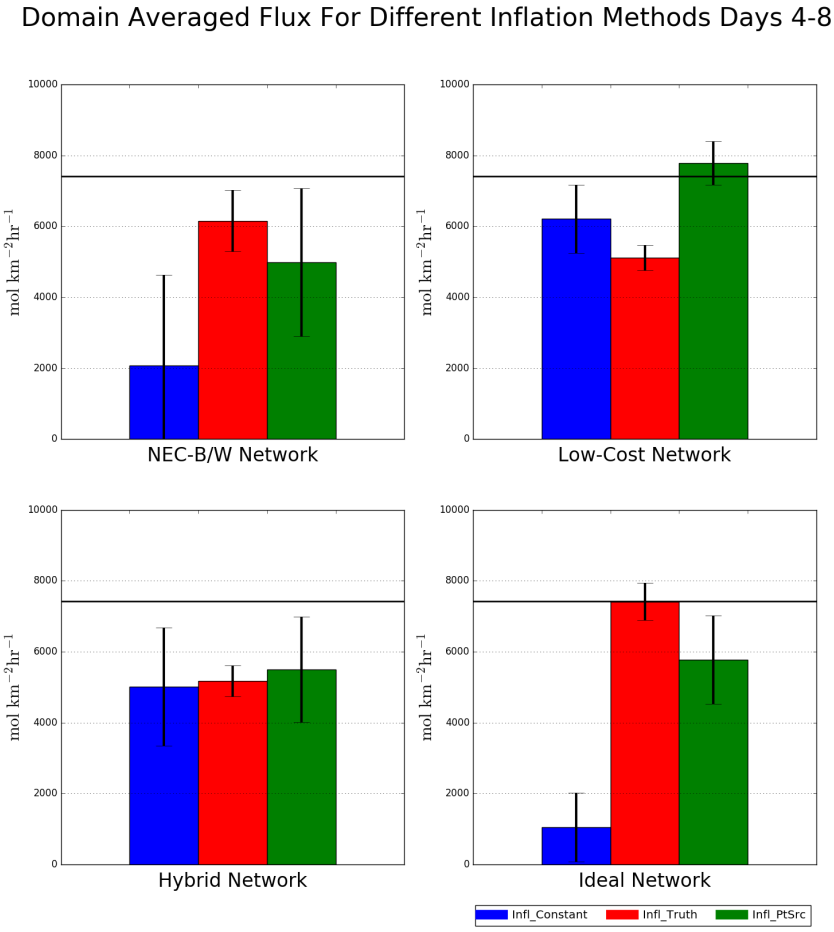

Figure 4.4. Domain d02 averaged surface CO_2 flux (bars) and standard deviation (error bars) in mol km^{-2} hr^{-1} for February 4 through February 8 2016 for each of the four observing networks (Top-Left: NEC-B/W; Top-Right: Low-Cost; Bottom-Left: Hybrid; Bottom-Right: Idealized) and for each of the three additive inflation methods (Blue: Infl_Constant; Red: Infl_Truth; Green: Infl_PtSrc) and the assumed true average flux in black.

Within this variability, the majority of these experiments approach the correct mean value for at least some of the assimilation cycles, but a more quantitative evaluation is needed to determine the overall performance of these estimations. Figure 4.4 shows the domain d02 average surface CO_2 flux (bars) and standard

deviation (error bars) for February 4 through February 8 2016 for each of the four observing networks and for each of the three additive inflation methods. The period of five days from February 4 through February 8 2016 was chosen to allow the system to spin up and approach an equilibrium, thus removing the first three days, and the last two days of the simulations were removed to exclude the aforementioned decrease observed across many of the experiments.

Table 4.2. Analysis mean and standard deviation over days 4-8 (February 4 – February 8, 2016) of the domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} for each of the four observing networks and the three additive inflation methods.

Analysis Mean and Standard Deviation Over Days 4-8
True Mean: 7418 Units: mol km^{-2} hr^{-1}

Network	Infl_Constant Mean	Infl Constant Std. Dev.	Infl_Truth Mean	Infl Truth Std. Dev.	Infl_PtSrc Mean	Infl PtSrc Std. Dev.
NEC-B/W	2081	2554	6152	866	4986	2082
Low-Cost	6209	965	5110	355	7786	609
Hybrid	5012	1667	5176	437	5500	1489
Ideal	1056	967	7414	528	5775	1251

For the two networks with only high accuracy observations, Infl_Constant performs quite poorly, with a mean value dramatically lower than the other two inflation methods, and a standard deviation that is comparable to or even larger than the mean itself. This is not nearly as dramatic in the Hybrid or Low-Cost network which incorporate 200 of the lower-accuracy observations, but they do still tend to have higher uncertainty and a lower mean than other configurations with the same observing network. Incorporating the point sources only, as is done in Infl_PtSrc, changes the results in all four observing networks, but only slightly in the Hybrid network. For the two networks with only high-accuracy observations, the mean value changes dramatically for the better, but the standard deviations either decrease as is

the case for NEC-B/W, or increase for the Ideal network. The best performer of these 12 experiments is unsurprisingly when using the idealized network with Infl_Truth. Not only is the mean value of this period within less than 0.1 % of the nature mean value, but the variations about the mean are less than 5%. This small uncertainty for Infl_Truth holds true across all the observing networks, and for all three inflation methods the Low-Cost network has the smallest standard deviations in the estimated domain mean surface fluxes of CO_2. Out of these twelve experiments, only two (Ideal-Infl_Truth and Low-Cost-Infl_PtSrc) have an estimated domain average flux that is within 10% of the assumed true mean value of 7418 mol km^{-2} hr^{-1}. The mean values over days 4-8 can vary from as low as 1055 mol km^{-2} hr^{-1} for the Ideal network with Infl_Constant (which is only approximately 14% of the true mean), to the two aforementioned cases which are the highest and the closest to the assumed truth. Table 4.2 has the mean and standard deviation for all 12 of these experiments. These results suggest that the system needs to have some sort of prior information on the spatial distribution and relative magnitude of uncertainty to properly estimate the fluxes.

In addition to the differences in the time series and mean estimated surface fluxes of CO_2, one can also evaluate the differences in the spatial distribution of the estimated fluxes between the different inflation techniques and observation networks. Figure 4.5 shows spatial plots for all 12 of these experiments, three inflation techniques for each of the four observing networks, and features the average surface flux over days 4-8 of the assimilation experiments. Immediately one can see the effect introducing prior information has on the flux estimation, when using Infl_Truth the

magnitude and pattern of the estimated fluxes look comparable to the assumed true fluxes from Fig. 4.1 regardless of observing network used. The other two inflation methods feature much more gradual gradients across the domain, with the exception of the point sources in Infl_PtSrc which are appropriately much larger than their surrounding pixels.

Rather than looking at the absolute values of the mean analysis flux, Fig. 4.6 shows the percentage difference between the estimated mean flux averaged over days 4-8 for each of the 12 experiments compared to the assumed true surface CO_2 flux. For the ideal network configurations, it's more apparent in these plots that Infl_Constant and Infl_PtSrc lead to large regions of underestimation, whereas with Infl_Truth, most of the region is only slightly over or underestimated compared to the true value, consistent with the domain averaged results shown earlier. Another thing to note is that across the other three observation networks, regardless of inflation method used, the Low-Cost network tends to have smaller errors in the urban core of the domain (the corridor running from Washington, DC northeast to Philadelphia, PA), consistent with the parts of the domain with the most of these observations. This area has a larger low bias (even outside of the immediate urban areas) in the Hybrid network, and this trend continues for the NEC-B/W network configuration. This is most easily visualized by looking at the ratio of red to blue in the panels on the left and right of Fig. 4.6.

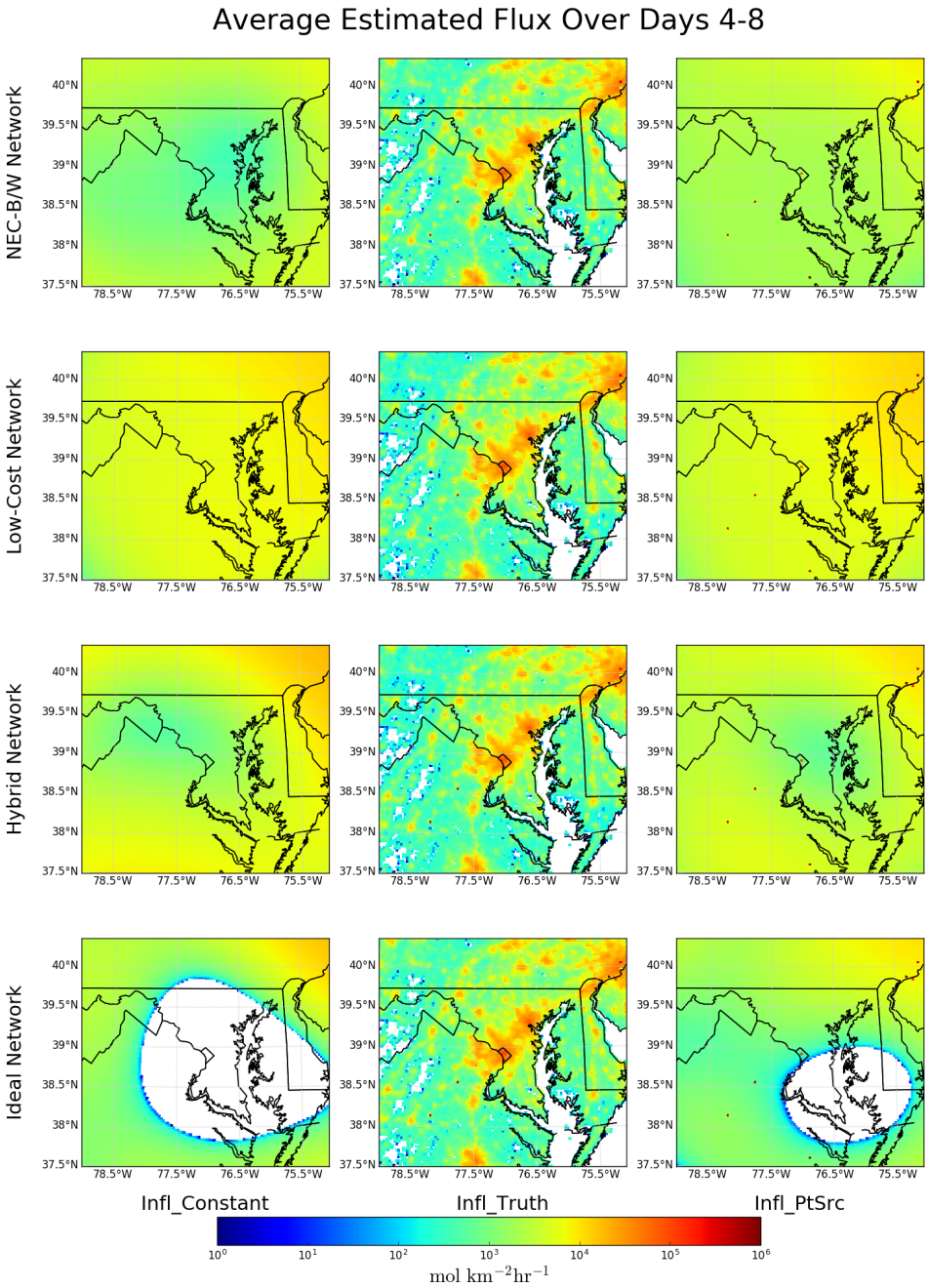

Figure 4.5. Spatial plots of the estimated surface CO_2 fluxes averaged over days 4-8 of each experiment for each of the four observation networks (four rows) and three inflation techniques (three columns). White areas are fluxes estimated to be at or below zero.

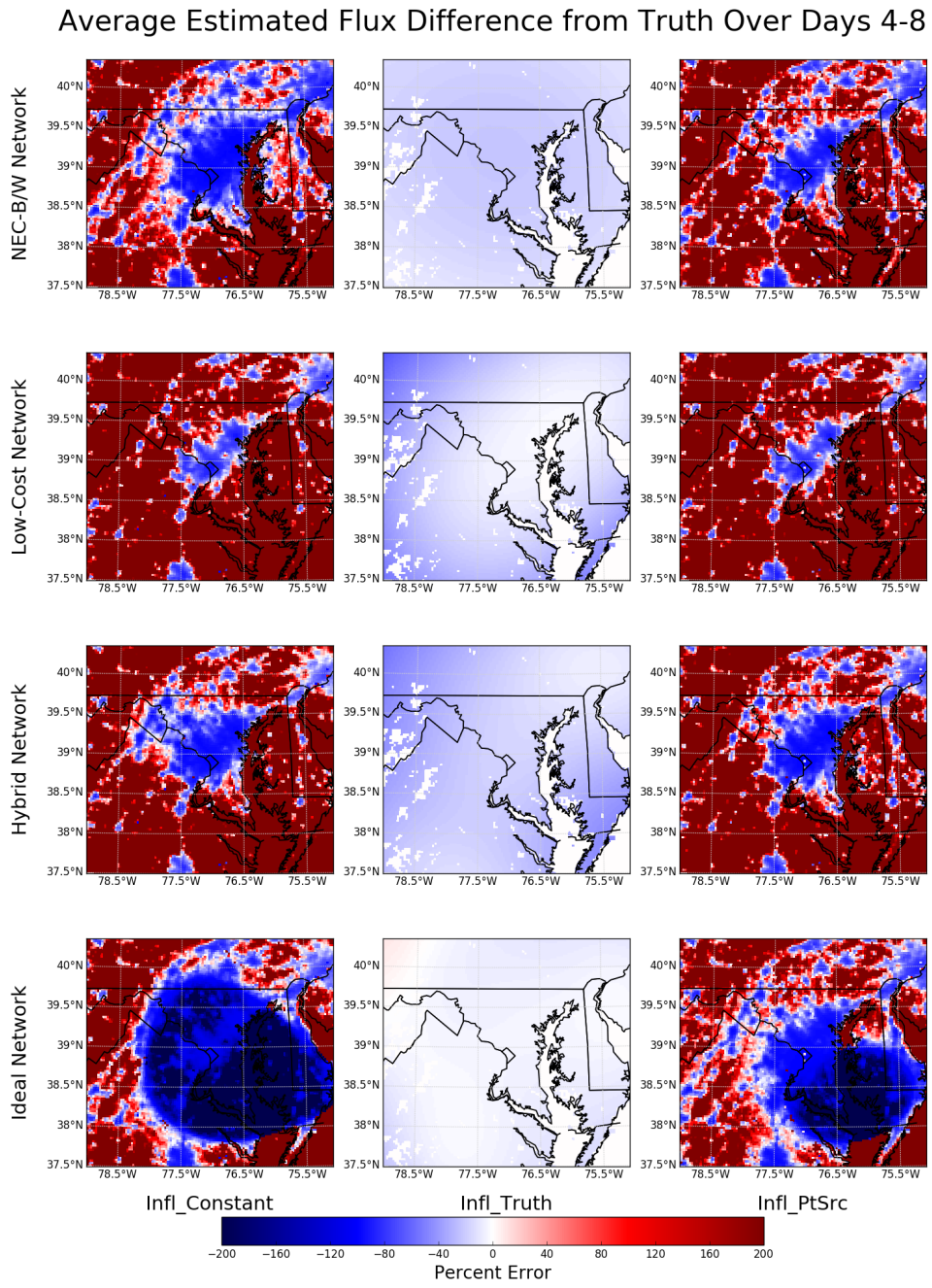

Figure 4.6. Spatial plots of the percent difference between the estimated surface CO_2 fluxes averaged over days 4-8 of each experiment and the assumed true fluxes for each of the four observation networks (four rows) and three inflation techniques (three columns).

4.4.2 Sensitivity to Observation Localization Radius

Section 4.3.3.3 described the observation localization radius and how it pertains to the LETKF data assimilation technique. This radius is related to the distance from a model grid point that an observation can be assimilated into the analysis. To evaluate the sensitivity of the estimated surface fluxes to this observation localization radius, both Infl_Constant and Infl_Truth inflation methods are used for a variety of localization radii for each of the four observing networks. Results using Infl_Constant are shown in Sect. 4.4.2.1 and using Infl_Truth in Sect. 4.4.2.2.

4.4.2.1 Using Infl_Constant

Rather than varying the additive inflation method used for a fixed localization radius as in Sect. 4.4.1, Figure 4.7 shows a time series of the domain average surface CO_2 flux for the nature run (black) as well as the estimated domain average surface CO_2 flux using the Infl_Constant additive inflation method for all experiments but the observation localization radius (each color represents a different localization radius) varies. Note that the 1 km radius is only used for the Ideal network, as it is the only one with sufficient observational density to work with only including observations within its grid cell. However, all four networks each include 50 km, 100 km, 150 km, and 200 km radii for direct comparisons of the observation networks. Similarly to what was discussed in Sect. 4.4.1, there are discernable differences in the results regardless of observing network depending on which localization radius is used.

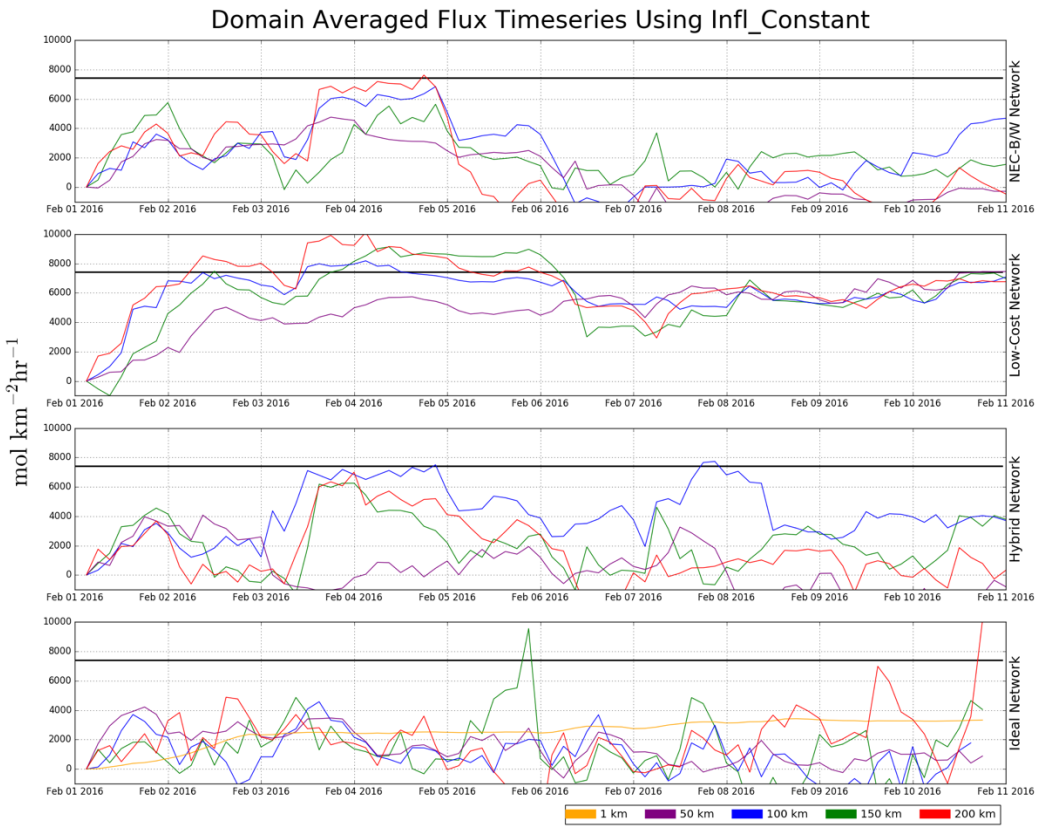

Figure 4.7. Time series of domain d02 average surface CO_2 flux in mol km^{-2} hr^{-1} using the Infl_Constant additive inflation method for each of the four observing networks (Top: NEC-B/W; Top-Middle: Low-Cost; Bottom-Middle: Hybrid; Bottom: Idealized) and for different observation localization radii in each color with the assumed true average flux in black.

For the Ideal network, the 1km radius features a relatively smooth curve of domain average surface flux, whereas the other localization radii all have large changes every few assimilation cycles. This is likely due to the fact that for the 1 km observation localization radius, only one observation is assimilated, whereas for larger radii, thousands of observations can be influencing each grid point. The 200 km radius experiment while behaving similarly to the others at first, towards the end actually approaches and even exceeds the domain averaged flux for the final cycle. For the other localization radii, and for all of the observing networks, the results,

while different, tend to follow a similar pattern in time for a given observing network, particularly the radii larger than 50 km. Generally, for each network, the four experiments tend to have correlated derivatives, meaning that when the flux in one increases, they all increase, and vice-versa. The biggest difference in the different experiments tends to be the mean value of the domain averaged fluxes: the mean value can approach the true mean value, or can be a fraction of the truth at any given analysis cycle depending on which localization radius is used. This is most clearly shown in Fig. 4.8 where again the bars represent the time mean of the domain averaged flux for days 4-8, and the associated error bars are the standard deviations of that temporal mean.

Of these 17 experiments using Infl_Constant, none of them have a mean value over days 4-8 that is within 10% of the assumed true value of 7418 mol km^{-2} hr^{-1}. All four of the experiments using the Low-Cost network perform the best, with the 150 and 200 km localization radii tests having the true mean within their 1-sigma standard deviations about their mean. Typically, the Low-Cost network experiments also feature the smallest standard deviations along with the largest mean values. With only 20 observations and Infl_Constant, the four NEC-B/W experiments are the worst performers overall, with domain mean fluxes that are largely negative during much of days 4-8, likely due to the lack of any sufficient physical constraint in the magnitude of the surface fluxes. The mean and standard deviation over days 4-8 for all 17 of these Infl_Constant experiments are available in Table 4.3.

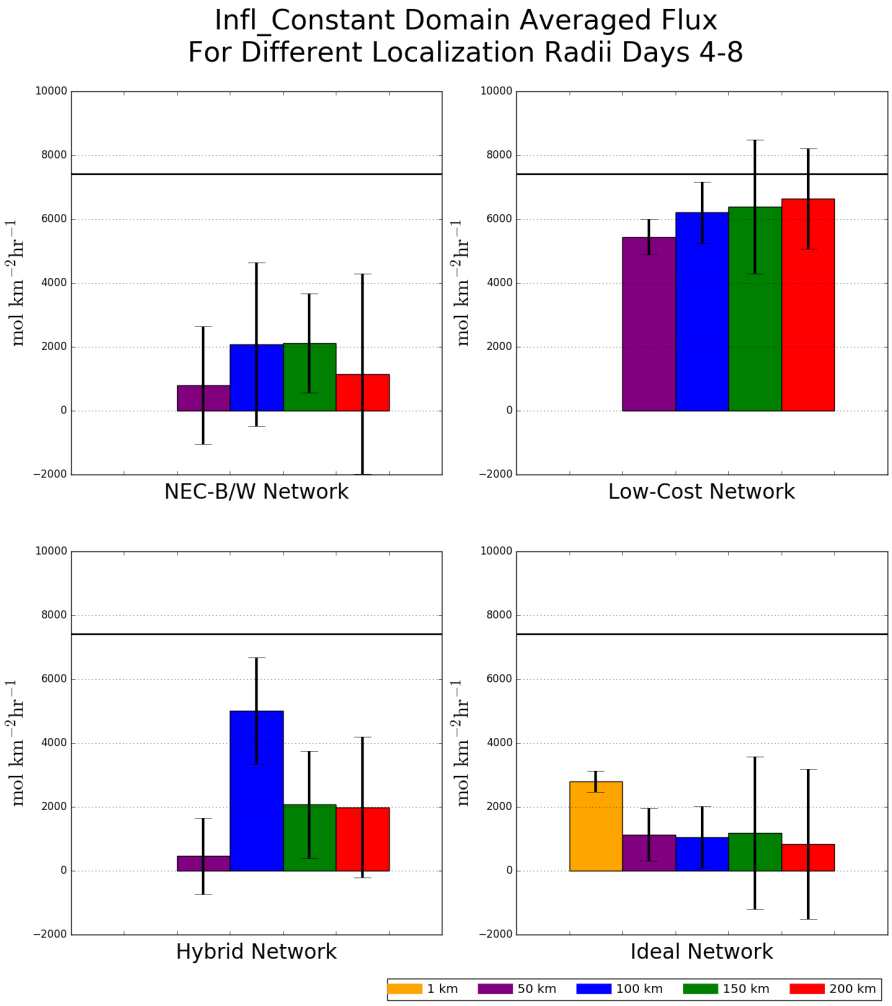

Figure 4.8. Domain d02 averaged surface CO_2 flux (bars) and standard deviation (error bars) in mol km^{-2} hr^{-1} for February 4 through February 8 2016 for each of the four observing networks (Top-Left: NEC-B/W; Top-Right: Low-Cost; Bottom-Left: Hybrid; Bottom-Right: Idealized) and for each of the observation localization radii used (Orange: 1 km; Purple: 50 km; Blue: 100 km; Green: 150 km; Red: 200 km) with Infl_Constant, and the assumed true average flux in black.

Table 4.3. Analysis mean and standard deviation over days 4-8 (February 4 – February 8, 2016) of the domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} for each of the four observing networks, and each observation localization radius using Infl_Constant.

Infl_Constant Analysis Mean and Standard Deviation Over Days 4-8 True Mean: 7418 Units: mol km^{-2} hr^{-1}				
Network	**NEC-B/W**	**Low-Cost**	**Hybrid**	**Ideal**
Mean - 1 km	N/A	N/A	N/A	2795
Std. Dev. - 1km	N/A	N/A	N/A	332
Mean - 50 km	795	5441	456	1134
Std. Dev. - 50 km	1842	555	1196	822
Mean - 100 km	2081	6209	5012	1056
Std. Dev. - 100 km	2554	965	1667	967
Mean - 150 km	2119	6388	2071	1182
Std. Dev. - 150 km	1551	2101	1679	2394
Mean - 200 km	1156	6639	1986	839
Std. Dev. - 200 km	3138	1578	2208	2350

In addition to variations in the domain averaged surface flux, each localization radius affects the spatial distribution of the analyzed CO_2 fluxes. In Fig. 4.9, it is apparent that the choice of localization radius can drastically affect the spatial pattern when using Infl_Constant regardless of observing network. One thing to note is that for the NEC-B/W network and the Low-Cost network, that using a small (50 km) observation localization radius can actually determine the general areas where emissions are largest (the urban regions) and smallest (rural areas to the west and southeast). Thus, it is plausible that areas of high and low emissions of CO_2 can be estimated or located without any prior spatial information introduced to the system. This concept will be evaluated further with additional experiments in a later section.

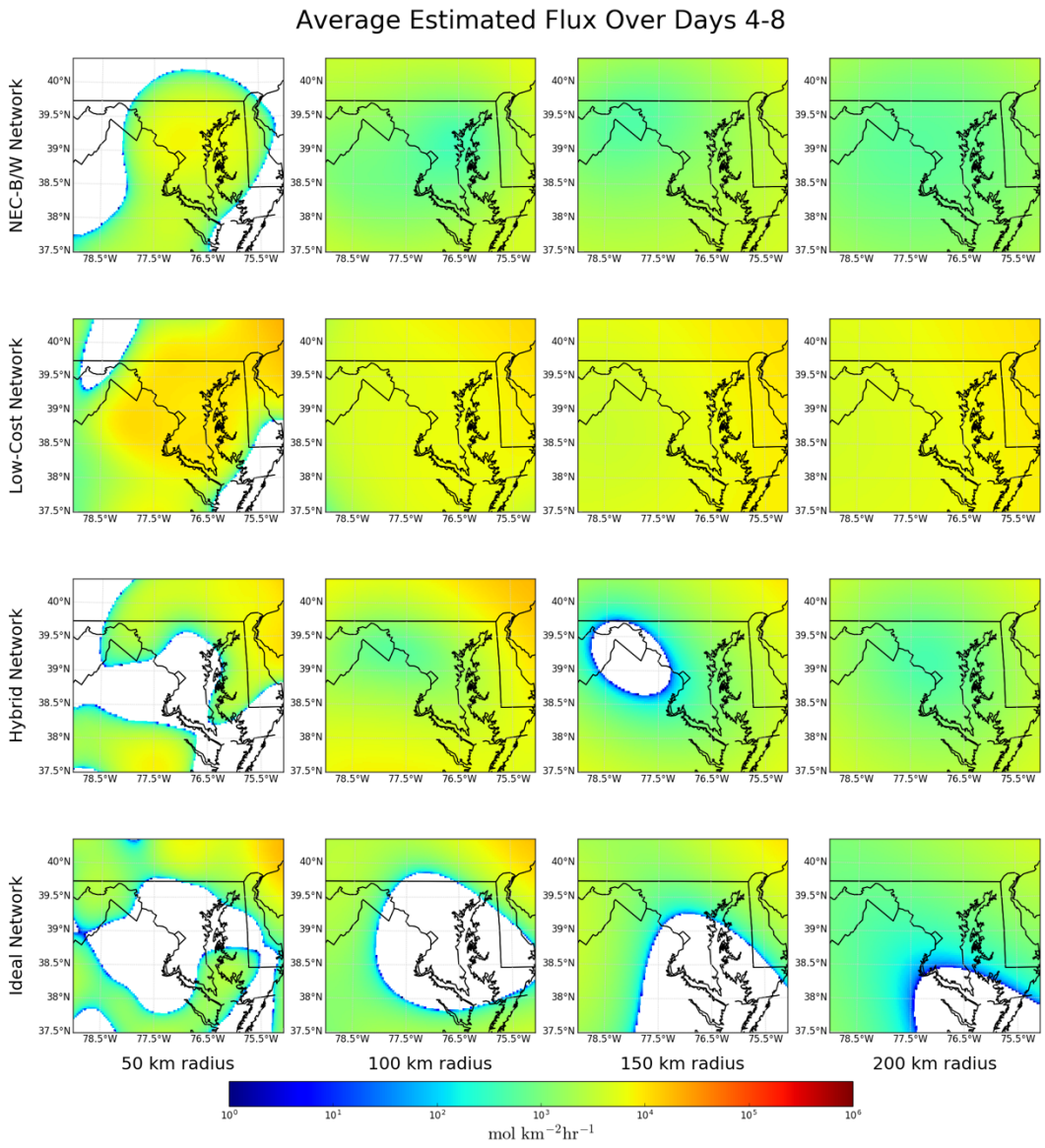

Figure 4.9. Spatial plots of the estimated surface CO_2 fluxes averaged over days 4-8 of each experiment for each of the four observation networks (four rows) and four observation localization radii (four columns) using the Infl_Constant additive inflation method. Areas in white are fluxes that are estimated to be at or below zero.

4.4.2.2 Using Infl_Truth

Section 4.4.1 showed that there is significant variability in the surface flux estimation depending on which additive inflation method is used. Section 4.4.2.1 used Infl_Constant in a variety of experiments varying the localization radius and

observation network used, but failed to find a satisfactory experiment that accurately estimated the true domain average surface flux of CO_2. In the following subsection, now we consider the same 17 cases from Sect. 4.4.2.1, but this time the experiments are redone with Infl_Truth as the additive inflation method used. Figure 4.10 shows a time series of the domain average surface CO_2 flux for the nature run (black) as well as the estimated domain average surface CO_2 flux now using the Infl_Truth additive inflation method for all experiments but again the observation localization radius (each color represents a different localization radius) varies. Again, all four networks have 50 km, 100 km, 150 km, and 200 km experiments, and the Ideal network has the additional 1 km experiment.

The most apparent difference between the results using Infl_Constant and Infl_Truth is that the curves are much smoother in time. While more than half of the experiments using Infl_Constant featured standard deviations in the domain averaged surface CO_2 flux when considering days 4-8 that are larger than 1000 mol km^{-2} hr^{-1}, none of the experiments using Infl_Truth have standard deviations that large. This is likely due to the fact that while over a domain average the perturbations added are the same, Infl_Truth adds perturbations that are a percentage of the assumed true fluxes rather than a constant mean value, meaning that for any given pixel the changes between cycles would be smaller in Infl_Truth. Just like with Infl_Constant, for any given observation network, the results are different for each localization radius, but all of the time series behave similarly, with the only substantial differences being the magnitude of the estimated surface fluxes. The general trend is, the larger the

observation localization radius and thus the more observations included in the analysis at each model grid point, the higher the domain averaged surface CO_2 flux is.

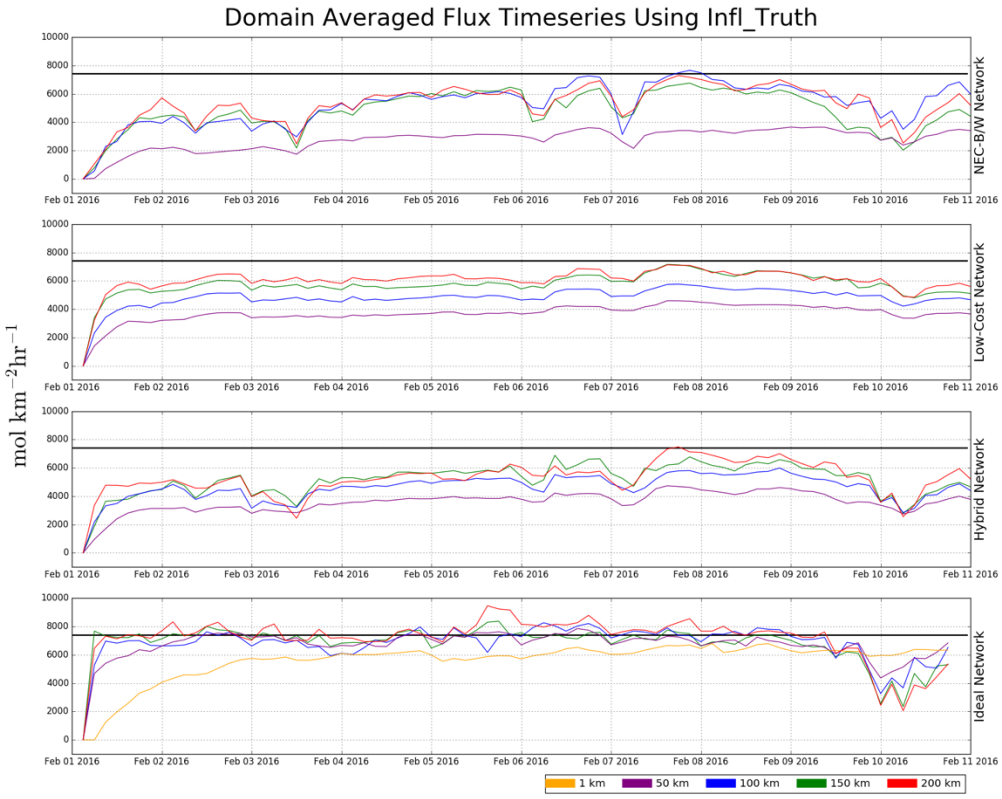

Figure 4.10. Time series of domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} using the Infl_Truth additive inflation method for each of the four observing networks (Top: NEC-B/W; Top-Middle: Low-Cost; Bottom-Middle: Hybrid; Bottom: Idealized) and for different observation localization radii in each color with the assumed true average flux in black.

With the assumed true domain averaged surface CO_2 flux again being a constant value of 7418 mol km^{-2} hr^{-1}, three of the experiments with the Ideal observing network have an estimated domain average surface CO_2 flux over days 4-8 within 5% of the true value, and another within 10% (Fig. 4.11). Unfortunately, none of the experiments from the other three observation networks can accurately estimate the total domain's average flux within 10%. Four of the experiments (two each using

the NEC-B/W network and the Low-Cost network) have estimates that are approximately 6200 mol km^{-2} hr^{-1}, but that is approximately 15% below the true estimated value. See Table 4.4 for the means and standard deviations for all of these experiments. While this would suggest that these observing networks cannot reliably estimate the surface CO_2 flux, these values are for a five-day average. When looking at the time series for each experiment, at least one localization radius for each of the four observation networks estimates a value that is very close to the true domain mean for one or more analysis cycles. In Chapter 3, we noted the possibility of synoptic meteorology playing a role in CO_2 transport and dispersion, and that future work needed to investigate metrics or methods to evaluate whether or not the certain synoptic situation was ideal for CO_2 flux estimation. Thus, it is plausible that for only including ideal analysis cycles that the surface CO_2 flux estimates would be closer to the assumed truth.

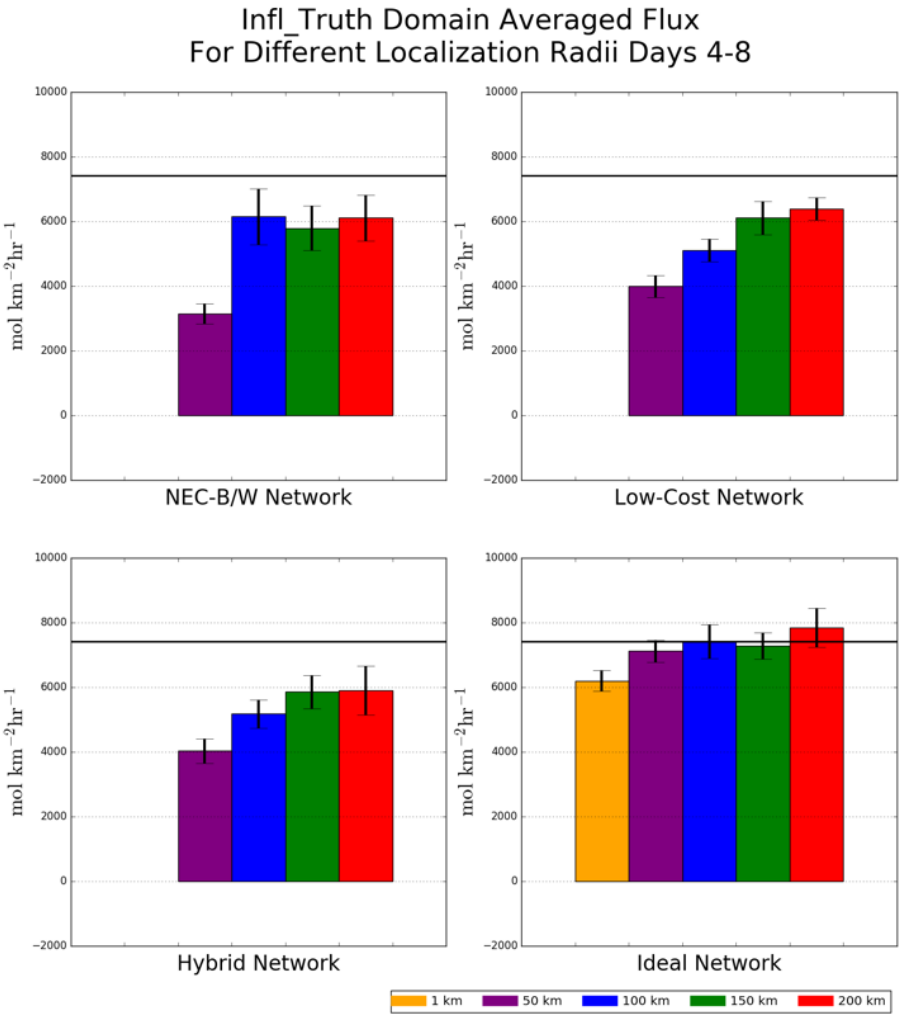

Figure 4.11. Domain d02 averaged surface CO_2 flux (bars) and standard deviation (error bars) in mol km^{-2} hr^{-1} for February 4 through February 8 2016 for each of the four observing networks (Top-Left: NEC-B/W; Top-Right: Low-Cost; Bottom-Left: Hybrid; Bottom-Right: Idealized) and for each of the observation localization radii used (Orange: 1 km; Purple: 50 km; Blue: 100 km; Green: 150 km; Red: 200 km) with Infl_Truth, and the assumed true average flux in black.

Table 4.4. Analysis mean and standard deviation over days 4-8 (February 4 – February 8, 2016) of the domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} for each of the four observing networks, and each observation localization radius using Infl_Truth.

Infl_Truth Analysis Mean and Standard Deviation Over Days 4-8 True Mean: 7418 Units: mol km^{-2} hr^{-1}				
Network	**NEC-B/W**	**Low-Cost**	**Hybrid**	**Ideal**
Mean - 1 km	N/A	N/A	N/A	6194
Std. Dev. - 1km	N/A	N/A	N/A	322
Mean - 50 km	3146	3992	4027	7123
Std. Dev. - 50 km	310	341	373	340
Mean - 100 km	6152	5110	5176	7414
Std. Dev. - 100 km	866	355	437	528
Mean - 150 km	5795	6114	5859	7275
Std. Dev. - 150 km	688	515	514	407
Mean - 200 km	6116	6392	5905	7843
Std. Dev. - 200 km	709	355	754	597

Because experiments using Infl_Truth provide spatial patterns that are the same as the assumed true emissions, one cannot easily discern differences in the spatial plotted fluxes for each experiment. However, Fig. 4.12 shows the percent error of each experiment's average flux over days 4-8 compared to the assumed true fluxes. Like with the Infl_Constant experiments, the localization radius makes a difference on the spatial distribution of the fluxes in the analysis. Because the observations for the three networks besides Ideal tend to be in the center of the domain rather than at the edges, for the 50km radius, there is a distinct ring around the urban center where fluxes are closer to the truth inside, and very low on the outside of this ring. For the larger radii, this pattern is not noticeable but one can still note how, generally, across the networks, when using Infl_Truth, a larger localization radius, and thus including more observations, reduces the error at each pixel across the domain.

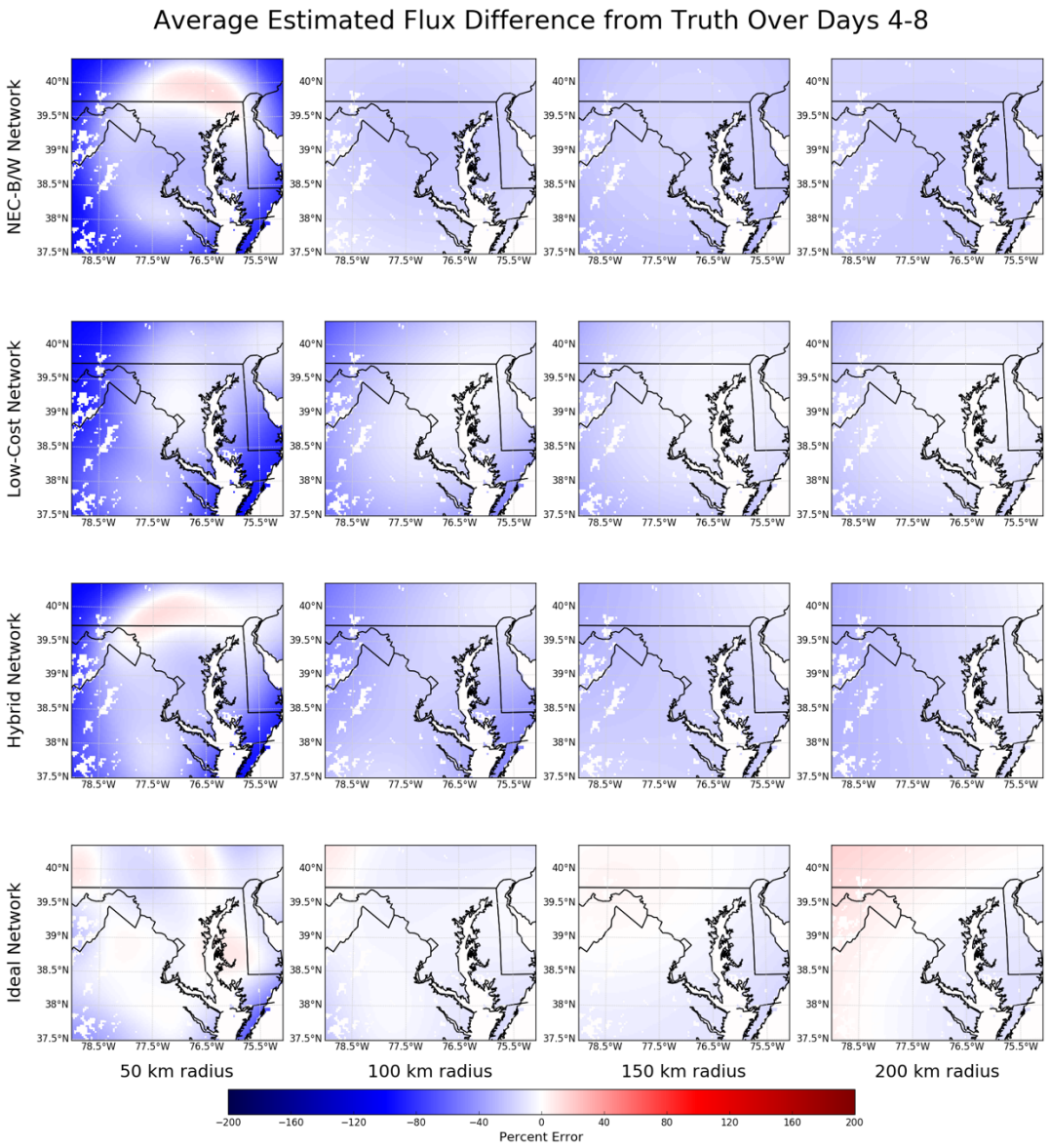

Figure 4.12. Spatial plots of the percent error of the estimated surface CO_2 fluxes averaged over days 4-8 compared to the assumed true fluxes from each experiment for each of the four observation networks (four rows) and four observation localization radii (four columns) using the Infl_Truth additive inflation method.

4.4.3 Additional Sensitivity Experiments

The previous two subsections focused on how the additive inflation methodology and observation localization radius affected the surface CO_2 flux estimation using LETKF. To briefly examine how other configuration choices may affect the analysis estimated flux, three additional experiments are performed. First, we can look at how doubling the perturbation size added as part of the additive inflation after the analysis affects the results. Additionally, the distribution of the observations as well as increasing the observation number by 50% are evaluated. For all three of these experiments, Infl_Truth is used as it provided the best overall results, and the lower accuracy observations (Obs_LC) are only used, but the observation localization radius differs for one of the three experiments.

4.4.3.1 Increasing Inflation Scaling Coefficient

In Sect. 4.3.3, the additive inflation ensemble perturbation method used in all of these experiments in this study is described. Within that section, a scaling coefficient that is multiplied to the randomly generated perturbations about a mean before the perturbations are added to each ensemble member is defined to be 0.1 for the OSSEs shown in this evaluation. However, this coefficient is simply a configurable option in the LETKF software, and thus, is simple to change for a sensitivity experiment. Figure 4.13 has both a time series and average over days 4-8 for the Low-Cost network with Infl_Truth and a 200 km observation localization radius, with the original experiment from Sect. 4.4.2.2 in red, and a new experiment

with the additive inflation coefficient doubled to 0.2 in blue. There are some notable differences observed in the time series of surface CO_2 flux, most notably that the initial increase in domain averaged flux is larger with the 0.2 coefficient versus the 0.1 experiment. The general pattern of the time series is the same, and the last few days are very similar to one another, but the 0.2 coefficient experiment is almost always higher, and thus closer to the true value, than the 0.1 experiment. For the mean value over days 4-8 of the experiments, the 0.2 coefficient run is a little more than 3% higher at 6603 mol km^{-2} hr^{-1} compared to 6392 mol km^{-2} hr^{-1} when averaged over the entire domain. The final thing of note is that, unsurprisingly since the perturbations are larger, the standard deviations are now larger with the larger coefficient, with an approximate increase of 22% from 355 to 444 mol km^{-2} hr^{-1}.

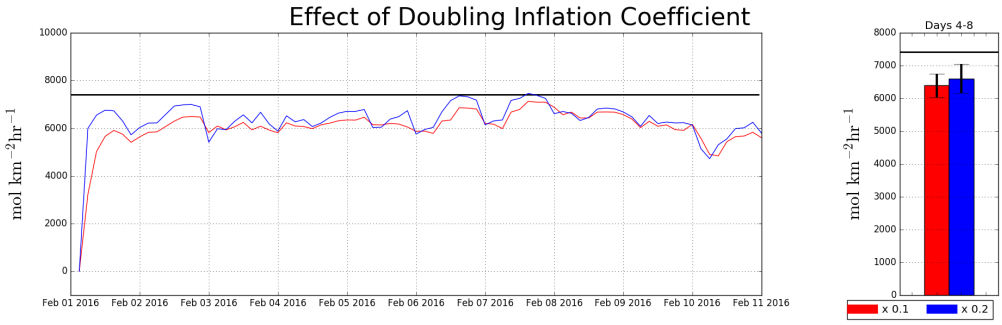

Figure 4.13. Time series as well as the mean (bars) and standard deviation (error bars) over days 4-8 of domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} using Infl_Truth and an observation localization radius of 200 km for the Low-Cost network for an additive inflation perturbation coefficient of 0.1 (red) and 0.2 (blue) with the assumed true flux in black.

4.4.3.2 Spatial Distribution of Observations

The Low-Cost network defined in Sect. 4.3.2 features 200 Obs_LC observations placed only in pixels that WRF-Chem designated as urban/developed areas. However, this puts a large concentration of observations in the urban center of the domain and leaves the outer edges with few or no observations. Instead, the same number of observations can be redistributed randomly about the domain and the results can be compared to see how much of a difference the observation network design makes when using an observation localization radius of 100 km. In Fig. 4.14, a time series as well as the day 4-8 mean and standard deviation for experiments using the Low-Cost network with Infl_Truth and a 100 km observation localization radius are shown, with the original experiment from Sect. 4.4.2.2 in red, and a new experiment with the observations randomly distributed in blue. The randomly distributed sensor network experiment takes longer to stabilize the domain averaged surface CO_2 flux. While the original Low-Cost network experiment starts to level off after one day, this doesn't occur until the end of the second day for the random observation location network. Additionally, the mean value for this random network over days 4-8 is over 8% lower than the original Low-Cost network, decreasing to 4711 mol km^{-2} hr^{-1} from 5110 mol km^{-2} hr^{-1}. This is likely due to the fact that in the original network, more of the observations were in urban areas and saw greater CO_2 enhancements, and thus, adjusted the fluxes higher than when there are more observations in the rural parts of the domain. The standard deviation of the domain averaged surface fluxes increases similarly as in the previous case, with a change

approaching 25% of the original standard deviation (albeit still under 500 mol km^{-2} hr^{-1}).

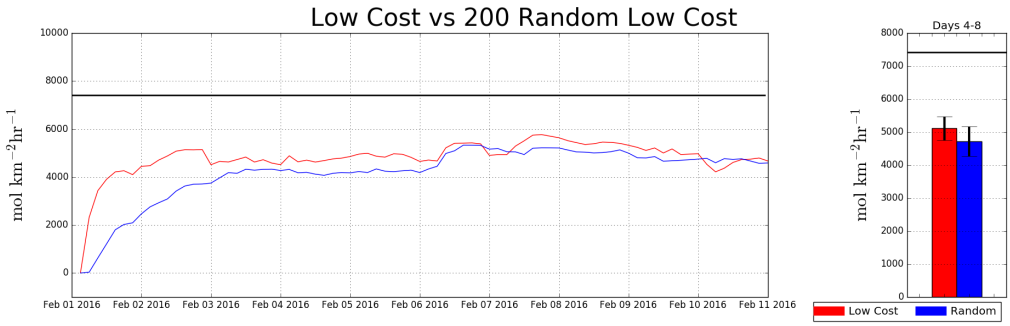

Figure 4.14. Time series as well as the mean (bars) and standard deviation (error bars) over days 4-8 of domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} using Infl_Truth and an observation localization radius of 100 km for the Low-Cost network (red) and a random redistribution of the observations (blue) with the assumed true flux in black.

4.4.3.3 200 Obs_LC vs 300 Obs_LC

Finally, while the four observation networks evaluated in the previous sections cover a range of the distribution and number of observations, they also tend to only cover the extremes. Rather than having 200 randomly distributed Obs_LC observations throughout the domain, we can see how the estimation of the surface CO_2 flux changes by increasing this number by 50% to 300. Figure 4.15 again features a time series and days 4-8 average of experiments using Infl_Truth and a 100 km observation localization radius, the previous figure's 200 random Obs_LC observations experiment is again in blue, and an experiment with 300 randomly distributed Obs_LC observations is in red. During the first day, both experiments behave very similarly and produce near identical results. However, for the next few days of the simulation, the experiment with 300 observations consistently estimates a

higher domain averaged surface CO_2 flux. From midway through February 6 until the morning of February 9, both experiments again produce similar results. Finally, while this feature is not seen in the 200 random observation experiment, the 300 random locations experiment has a sudden drop in domain mean flux, consistent with many of the other experiments and likely is related to certain unfavorable meteorological conditions in part of the domain. However, continuing with the metric of the surface flux averaged over days 4-8, increasing the number of observations increases the flux estimate by about 7% and decreases the standard deviation by about 35%.

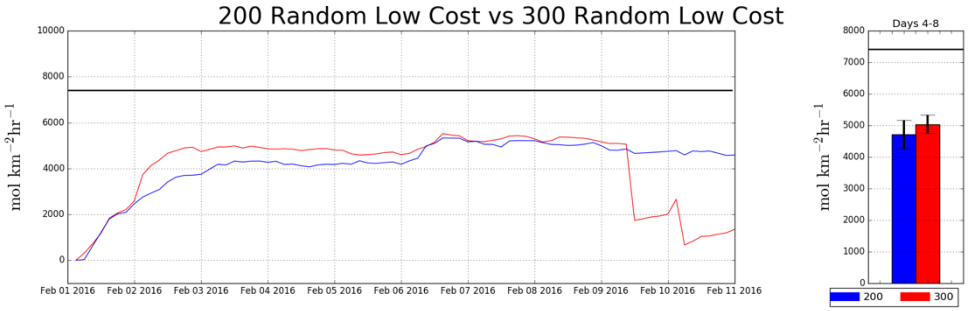

Figure 4.15. Time series as well as the mean (bars) and standard deviation (error bars) over days 4-8 of domain d02 averaged surface CO_2 flux in mol km^{-2} hr^{-1} using Infl_Truth and an observation localization radius of 100 km when using 200 randomly distributed Obs_LC observations (blue) and when using 300 randomly distributed Obs_LC observations (red) with the assumed true flux in black.

4.5 Discussion

4.5.1 Urban Center Performance

Section 4.4 focused on the evaluation of the performance of these OSSEs based on their ability to estimate the surface CO_2 flux averaged across the entire domain of interest and over days 4-8 of the experiment. With this criterion as the metric of success, only the Ideal network (for multiple localization radii) using

Infl_Truth was able to successfully estimate the CO_2 flux. Now we consider a situation where rather than estimating the CO_2 flux for the entire WRF (or other model) domain, only a certain subset of the domain is the area of interest for flux estimation. For example, the domain chosen is centered over the Baltimore, MD and Washington, DC metropolitan areas, and thus a goal could be just to reproduce the surface CO_2 fluxes from these urban areas. Figure 4.1 again shows the assumed true fluxes in domain d02, except now note that there is a box drawn around Baltimore and Washington that is 105 km x 105 km square (35 x 35 grid points). Now the experiments from Section 4.4 can be reevaluated by comparing their estimates in this urban box rather than the entire WRF domain.

When reevaluating the 12 experiments from Sect. 4.4.1 where the localization radius is constant at 100 km, but the additive inflation technique changes for the four observation networks, one can see immediately how much better Infl_Truth performs than the other two methods in the urban center. Figure 4.16 shows the time series of the urban area averaged surface CO_2 flux for these 12 experiments and the mean and standard deviations over days 4-8 are in Fig. 4.17. Infl_Truth has almost always the largest (and thus most correct) urban area averaged surface CO_2 flux, and Infl_Constant the lowest, which, depending on observation network, is frequently estimating a net sink of CO_2 in the urban center, completely different from the true relative maximum in the region for the assumed true emissions. When averaging over days 4-8 of the experiments, Ideal-Infl_Truth and Low-Cost-Infl_Truth both have estimates of the urban area averaged flux within 6% of the assumed true value. With the other two networks, using Infl_Truth can still cause an underestimation by as

large as 25% (from NEC-B/W). Using Infl_Constant, because there is no spatial constraint, the performance in the urban center is worse than for the total domain, with a net sink in the Ideal case, and a maximum average flux estimation that is less than 31% of the assumed true value. Using Infl_PtSrc fared better than Infl_Constant, but the urban area averaged CO_2 flux failed to get to within 25% of the true value.

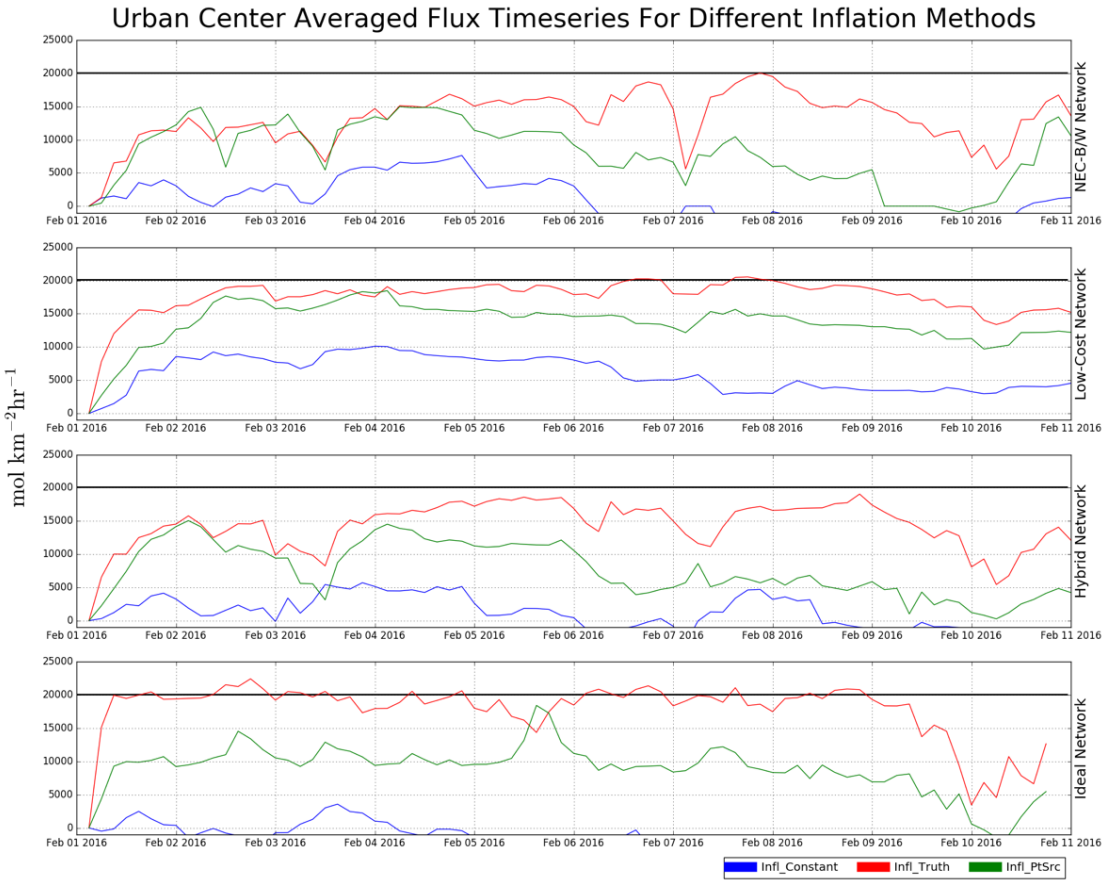

Figure 4.16. Time series of the urban region averaged surface CO_2 flux in mol km^{-2} hr^{-1} for each of the four observing networks (Top: NEC-B/W; Top-Middle: Low-Cost; Bottom-Middle: Hybrid; Bottom: Idealized) and for each of the three additive inflation methods (Blue: Infl_Constant; Red: Infl_Truth; Green: Infl_PtSrc) and the assumed true average flux in black.

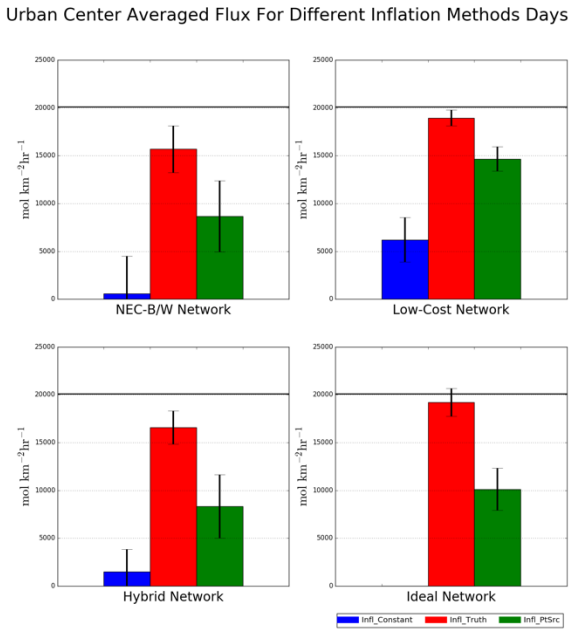

Figure 4.17. Urban region averaged surface CO_2 flux (bars) and standard deviation (error bars) in mol km^{-2} hr^{-1} for February 4 through February 8 2016 for each of the four observing networks (Top-Left: NEC-B/W; Top-Right: Low-Cost; Bottom-Left: Hybrid; Bottom-Right: Idealized) and for each of the three additive inflation methods (Blue: Infl_Constant; Red: Infl_Truth; Green: Infl_PtSrc) and the assumed true average flux in black.

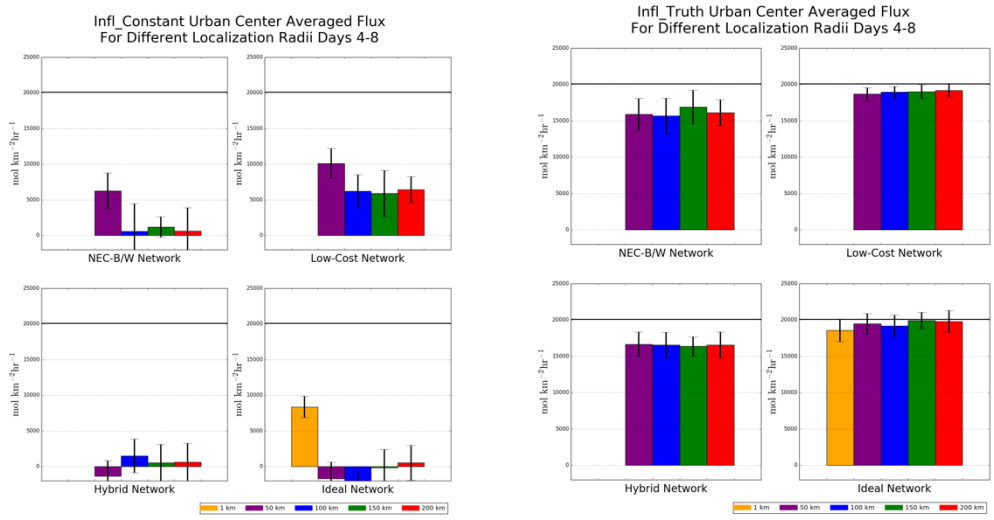

Figure 4.18. Urban region averaged surface CO_2 flux (bars) and standard deviation (error bars) in mol km^{-2} hr^{-1} for February 4 through February 8 2016 for each of the four observing networks, for Infl_Constant (left two columns) and Infl_Truth (right two columns), and for multiple observation localization radii (colors).

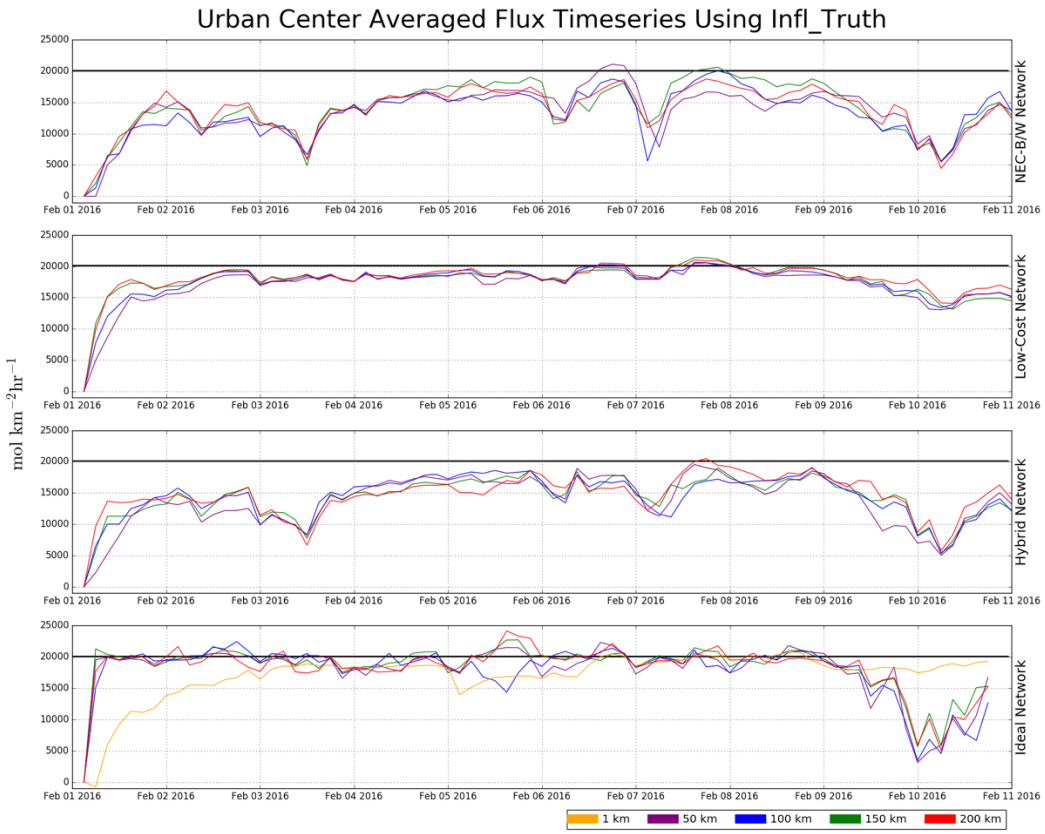

Figure 4.19. Time series of the urban region averaged surface CO_2 flux in mol km^{-2} hr^{-1} using the Infl_Truth additive inflation method for each of the four observing networks (Top: NEC-B/W; Top-Middle: Low-Cost; Bottom-Middle: Hybrid; Bottom: Idealized) and for different observation localization radii in each color with the assumed true average flux in black.

The impact that varying the observation localization radius has on the urban region averaged surface CO_2 flux depends greatly on which additive inflation method is used. Figure 4.18 features the mean and standard deviation of each localization radius for both Infl_Constant (the left panels) and Infl_Truth (the right panels) for all four observing networks. For Infl_Constant, none of the localization radii for any of the observing networks come close to properly estimating the urban area's average flux. For the networks with the largest number of observations (Hybrid and Ideal), some of the configurations cause the system to make the urban area a net sink of CO_2,

consistent with what was observed earlier, but is not limited to just one localization radius. The best performer with Infl_Constant (Low-Cost – 50 km) estimates a value that is approximately only half of the true urban average flux. When using Infl_Truth, the localization radius does not significantly impact the results in the urban area. For the Low-Cost and Hybrid networks, the difference between the highest of the four estimates of the urban mean flux averaged over days 4-8 and the lowest estimated value is less than 500 mol km^{-2} hr^{-1} for the Low-Cost network and less than 300 mol km^{-2} hr^{-1} for the Hybrid network. The differences between the experiments are larger for the NEC-B/W and Ideal observing networks, but there is still less than a 10% difference between the highest estimate and the lowest estimate. One final thing to note about the Infl_Truth experiments is that while for the entire domain generally the larger the localization radius, the higher the flux estimate, this is not true for the urban region, where in fact the largest localization radius only provides the highest flux for the NEC-B/W network (and the smallest localization radius tested provides the highest flux for the Hybrid network).

When averaging over the entire WRF domain d02, only the Ideal network can estimate the surface CO_2 flux within 5% or even 10%, as described in Sect. 4.4.2.2. However, when just considering this central urban region covering Baltimore and Washington, the results are more promising. In fact, the Low-Cost sensor network is able to estimate the urban area's mean flux within 5% when using an observation localization radius of 200 km. The other three localization radii produce results that are within 10% of the mean value. Both the NEC-B/W network and the Hybrid network fail to reach this threshold, with estimated mean values that are all

approximately 80% of the true average flux. The 1 km Ideal experiment estimates the urban region's average flux to within 10% of the true value, but all of the other four experiments are within 5% of the truth. All of the means and standard deviations of the urban region averaged surface flux using Infl_Truth are presented in Table 4.5. Again, as with the domain averaged estimates in Sect. 4.4.2.2, while the five-day averages are too low across the board, when looking at the time series plots (Fig. 4.19), each observation network has an ensemble configuration that can closely estimate the true urban area averaged surface CO_2 flux for one or more analysis cycles.

Table 4.5. Analysis mean and standard deviation over days 4-8 (February 4 – February 8, 2016) of the urban region averaged surface CO_2 flux in mol km^{-2} hr^{-1} for each of the four observing networks, and each observation localization radius using Infl_Truth.

Infl_Truth Analysis Mean and Standard Deviation Over Days 4-8 True Mean for Urban Area: 20114 Units: mol km^{-2} hr^{-1}				
Network	**NEC-B/W**	**Low-Cost**	**Hybrid**	**Ideal**
Mean - 1 km	N/A	N/A	N/A	18573
Std. Dev. - 1km	N/A	N/A	N/A	1586
Mean - 50 km	15912	18691	16666	19473
Std. Dev. - 50 km	2173	887	1683	1431
Mean - 100 km	15692	18952	16574	19202
Std. Dev. - 100 km	2453	822	1741	1468
Mean - 150 km	16923	19022	16374	19916
Std. Dev. - 150 km	2333	951	1340	1161
Mean - 200 km	16123	19174	16579	19814
Std. Dev. - 200 km	1791	887	1791	1495

4.5.2 Correlations of Truth vs Analysis

Section 4.4.1 showed that there is significant difference in the analyzed CO_2 flux depending on which additive inflation method is used. Without introducing any spatial information into the system (using Infl_Constant), generally the analysis is unable to estimate correctly the domain averaged surface flux. Fig 4.20 shows scatter plots comparing the analysis flux averaged over days 4-8 (y) with the assumed true flux (x) for each pixel in domain d02 for each of the four observing networks and for four localization radii using Infl_Constant and Fig. 4.21 is the same except for experiments using Infl_Truth. The comparison between the scatter plots on these two figures show how the introduction of spatial information into the ensemble perturbations and thus, into the CO_2 flux analysis improves the correlation between the analysis and the truth at each grid point. For the Infl_Constant cases the R^2 values are all virtually zero, with no linear correlation. The scatter plots support this, with almost all of the pixels in the analysis being near the mean value, with the true fluxes having substantial variability in their magnitude from pixel to pixel (from 0 to 10^7 mol km^{-2} hr^{-1}). Conversely, when using Infl_Truth, the correlations are quite good, with all of the experiments using a localization radius of 100 km or above having R^2 values of nearly 1. Interestingly, while one may suspect that the consistent underestimation of the domain averaged CO_2 flux was due to an underestimation of the point source (or the large pixels), the scatter plots do not support this hypothesis. For the small localization radius (50 km), the underestimation is most pronounced at values near the mean of the domain, but for all experiments there is a consistent low bias of the surface flux across all pixels.

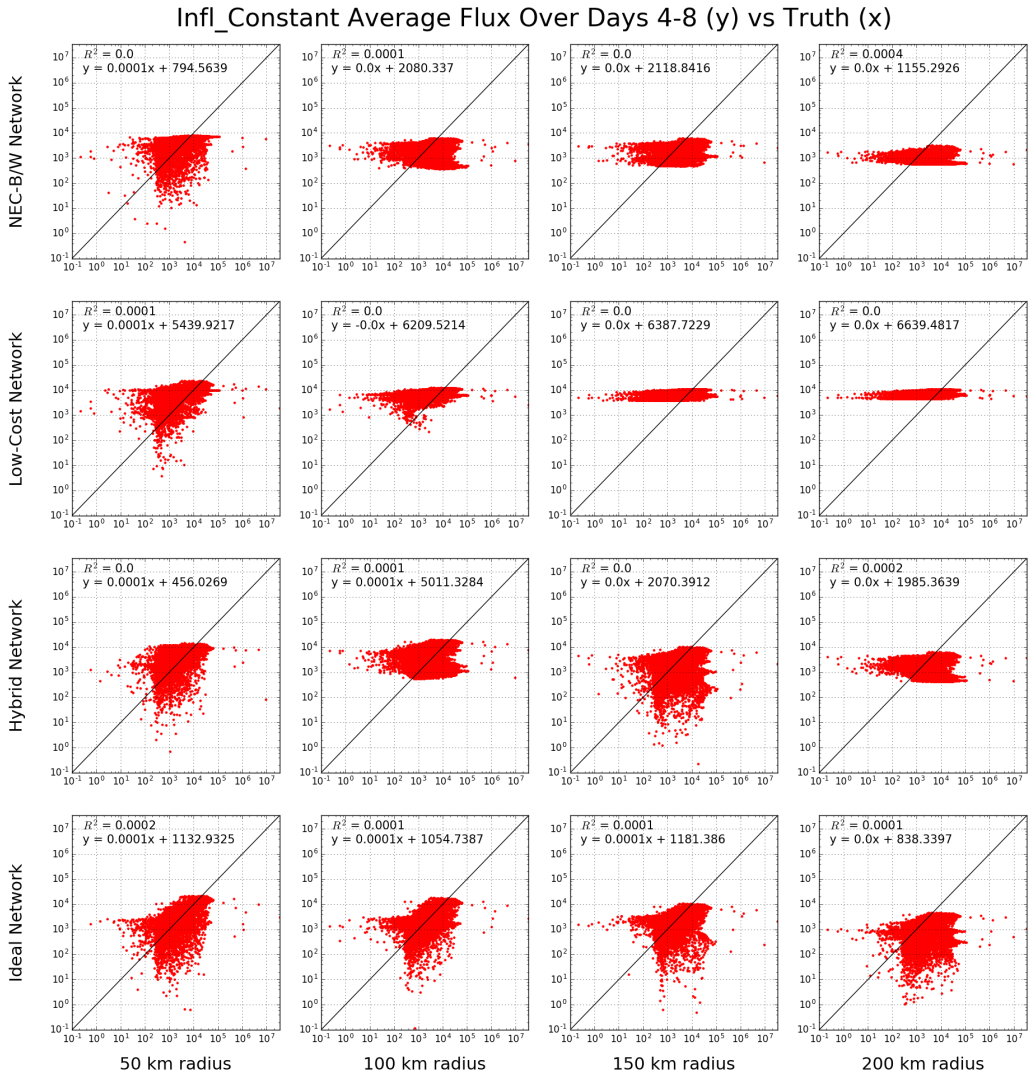

Figure 4.20. Scatter plots of the assumed true flux (x) and the analysis flux averaged over days 4-8 (y) for each model pixel for each observation network (the four rows) and for four observation localization radii (the four columns) using Infl_Constant (units are mol km^{-2} hr^{-1}).

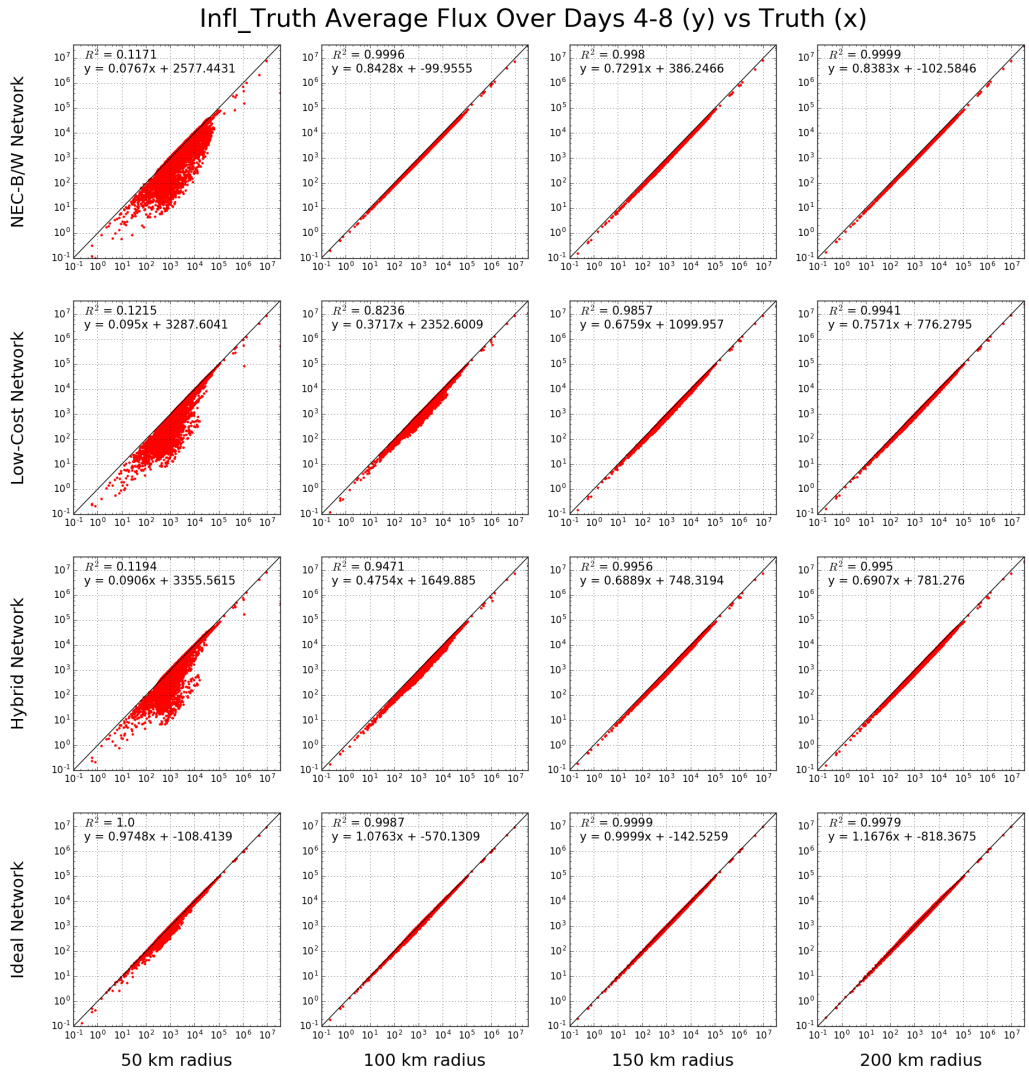

Figure 4.21. Scatter plots of the assumed true flux (x) and the analysis flux averaged over days 4-8 (y) for each model pixel for each observation network (the four rows) and for four observation localization radii (the four columns) using Infl_Truth (units are mol km^{-2} hr^{-1}).

4.5.3 Variability in the Analysis

The timeseries plots in Sect. 4.4 show variability across analysis times regardless of observing network, localization radius, or inflation technique used. Chapter 3 introduced this notion of how synoptic meteorology can play a role in the

quality of predicted CO_2, and this suggests that it may also play a role in the quality of the analysis of surface CO_2 fluxes. While the metric throughout most of this chapter is to take the average of days 4-8 and compare it to the assumed true flux, we can also consider the "best" analysis, meaning the analysis cycle where the domain averaged flux is closest to the true domain average, for each observing network and localization radius using the Infl_Truth method. Fig. 4.22 features spatial plots of the percent error of the analysis for each experiment's "best" analysis, including the valid time for this analysis and the mean difference from the truth.

Recall from Sect. 4.4 that the domain averaged true CO_2 flux is approximately 7418 mol km^{-2} hr^{-1}. Depending on observation localization radius, the domain averaged flux can be within 4% for the Low-Cost network, 1% for the NEC-B/W and the Hybrid networks, and within 0.05 for the ideal network (with a mean error of 3 or 4 mol km^{-2} hr^{-1} for the 100 and 150 km radii). These numbers suggest that when the system is best situated for estimating the flux, that the higher accuracy observations do perform better than the lower accuracy observation network. Additionally, all of these times save for the 50 km Ideal best case are between 0 UTC February 7 and 0 UTC February 9, supporting the hypothesis from Chapter 3 that there are periods where the overarching meteorological conditions allow for the model to estimate CO_2 better, and that this also may apply to flux estimation techniques. This suggests that further work needs to be done to quantify what periods or data should be included in an analysis system to get the optimum estimation of surface CO_2 fluxes.

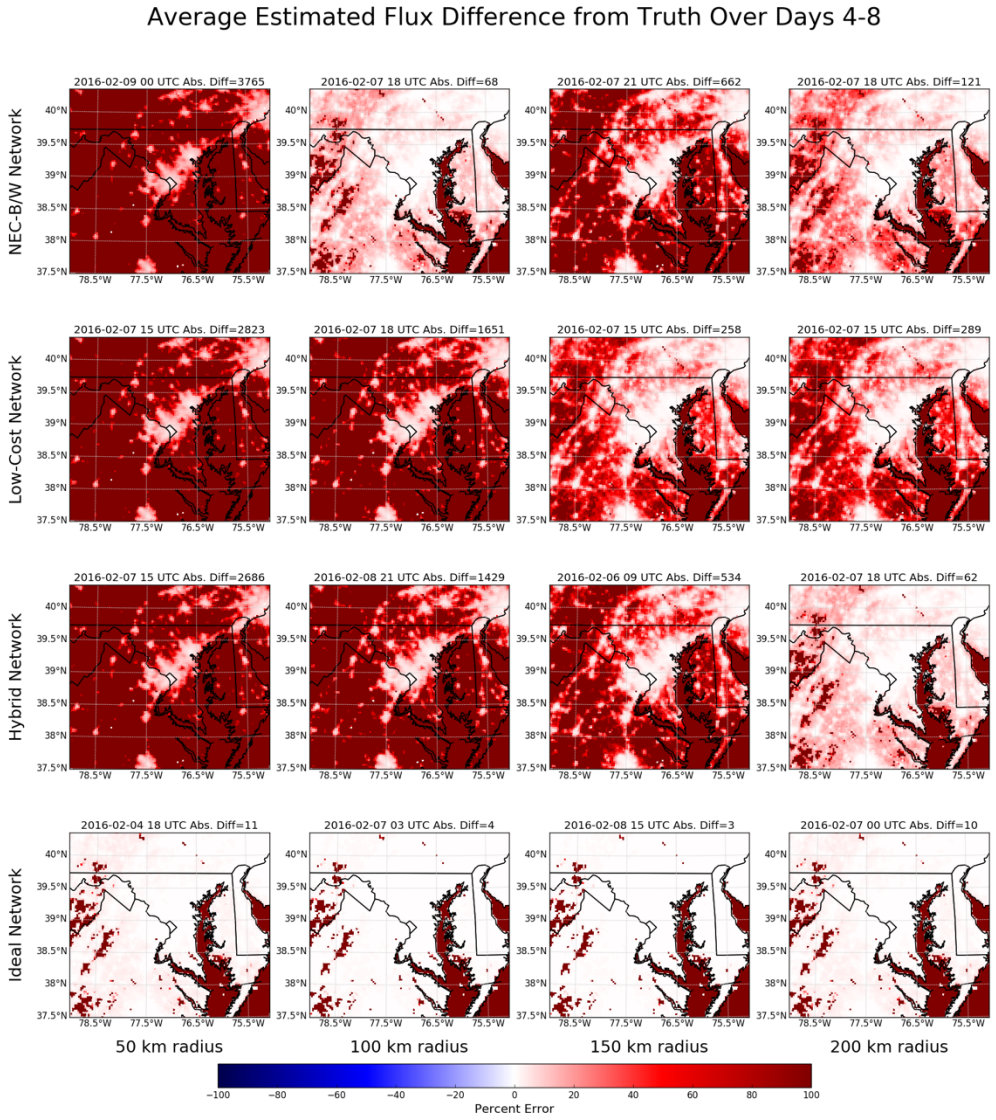

Figure 4.22. Spatial plots of the percent error of the "best analysis" for each experiment, meaning the one with the lowest mean absolute error compared to the true domain mean flux for each observation network (the four rows) and for four observation localization radii (the four columns) using Infl_Truth. The valid time of this optimum analysis is featured in the subplot title as well as this mean absolute error in units of mol km^{-2} hr^{-1}.

4.5.4 Estimating Spatial Pattern Using A Small Localization Radius

While the experiment using Infl_Constant and a 1 km localization radius for the Ideal network fails to get the correct mean flux for both the entire domain and just the urban center, Fig. 4.23 reveals something interesting. A high accuracy (Obs_HA) observation at every model grid point, and limiting the LETKF to only assimilate one observation at each point during each analysis cycle, allow for the general spatial pattern of the true fluxes to show up without any prior spatial information introduced into the data assimilation system, either through a forecasted prior flux or through additive inflation. The approximate regions of the Philadelphia, Baltimore, and Washington metropolitan areas are easily discernable in the surface CO_2 flux averaged over days 4-8 of the simulation. Additionally, all of the smaller cities and towns throughout the domain (as well as the rural areas) are also able to be picked up in this experiment, with magnitudes that are approximately consistent with the assumed true fluxes. However, the spatial means are too low compared to the truth, likely due to the fact that the point sources, which are much larger than the rest of the pixels and contribute significantly to the domain total, are not well resolved in this experiment.

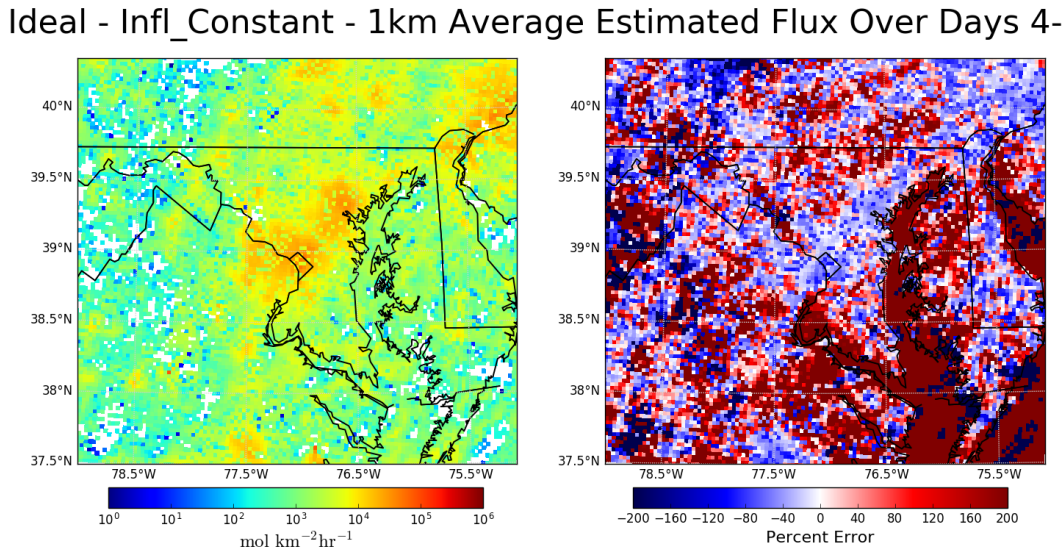

Figure 4.23. Spatial plots of the surface CO_2 flux estimated with the Ideal observing network, the Infl_Constant additive inflation method, and using a 1 km localization radius, averaged over days 4-8 of the experiment. Left panel shows the estimated flux and the right panel is the percent difference each pixel is from the assumed true fluxes.

This Ideal network of 13,320 Obs_HA observations is impractical both from a financial standpoint, but also logistical with the installation of that many observing locations in such a small area. Instead, we can consider a grid of 840 Obs_LC observations, each 12km apart from one another. Not only is the cost of each individual observing point much cheaper, but there is also an order of magnitude fewer observations. Fig. 4.24 shows that by decreasing the observational density but keeping them regularly spaced, and increasing the observation localization radius to 10 km, the system can still predict where the relative minimum and maximum fluxes should be. Just like with the 1km Ideal case, the domain averaged flux is approximately half of the true value, but without any spatial information introduced (aka using Infl_Constant), 840 Obs_LC locations regularly spaced can differentiate

between the large urban areas, the smaller cities, and the rural areas in the domain and their associated relative fluxes.

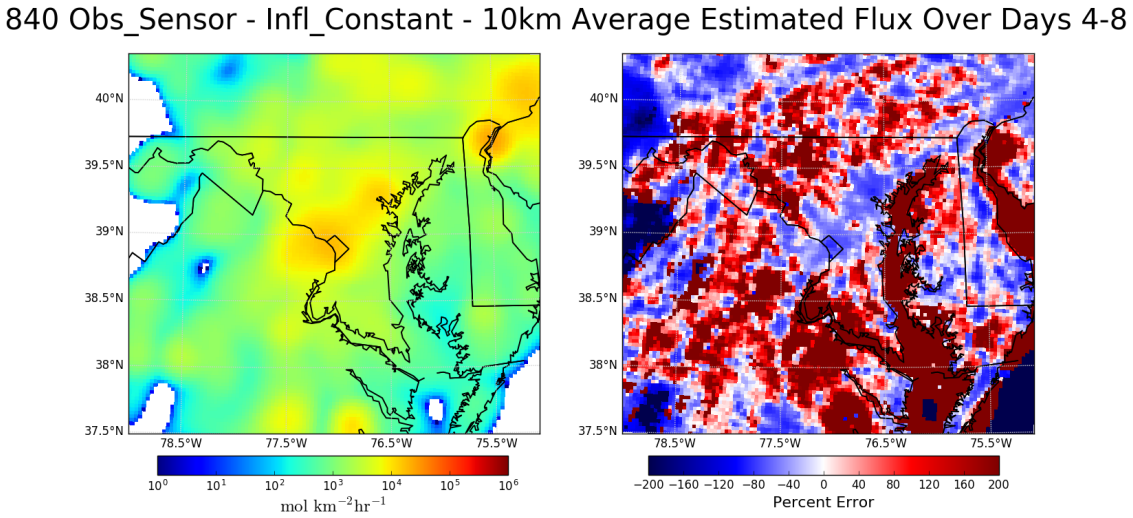

Figure 4.24. Spatial plots of the surface CO_2 flux estimated with 840 Obs_LC observing locations in a grid with spacing of 12 km, the Infl_Constant additive inflation method, and using a 10 km localization radius, averaged over days 4-8 of the experiment. Left panel shows the estimated flux and the right panel is the percent difference each pixel is from the assumed true fluxes.

4.5.5 Atmospheric CO_2 Analysis

While the focus of these OSSEs is to estimate the surface CO_2 flux for an urban region, one advantage that using data assimilation techniques such as LETKF has over inverse modeling, is that in addition to surface flux estimation, an analysis of the 3D CO_2 field is also generated at each cycle. In Fig. 4.25, the nature run domain averaged CO_2 is plotted in the thick black line, and each data assimilation experiment's domain average CO_2 is also plotted (blue lines using Infl_Constant, red lines using Infl_Truth, and green lines using Infl_PtSrc). Finally, a control run, meaning no data assimilation is performed, and thus the domain d02 surface CO_2 flux is zero for the entire experiment, is also plotted in dark gray. These plots show the

strong dependence that meteorology has on the observed CO_2 for a fixed flux (in the nature run) and that the contribution from the inner domain (the enhancement or the difference between the gray and black lines) can vary significantly in time due to the meteorology. Unsurprisingly, overall it seems that the experiments using Infl_Truth (red) provide the time series of CO_2 that most accurately represent the true diurnal and synoptic variability of the domain averaged CO_2. One thing of interest is that while most of the experiments underestimate the CO_2 concentrations, consistent with the overall underestimation of the surface flux, the experiments for the Low-Cost network using both Infl_Constant and Infl_PtSrc consistently overestimate the CO_2 concentrations averaged over the domain. These same experiments do not overestimate the surface CO_2 flux, so it's possible that this overestimation has to do with the incorrect distribution of the fluxes, allowing for erroneously high CO_2 concentrations in much of the domain. One final thing to note, is that Sect. 4.5.3 showed which analysis cycle had the mean value closes to the truth, and that this was often in the early afternoon on February 7 or February 8, which the CO_2 timeseries in Fig. 4.25 show that the days of February 6-8 have the largest diurnal variability in CO_2 during this 10-day period., further suggesting that meteorological conditions can affect the performance of the flux estimation system.

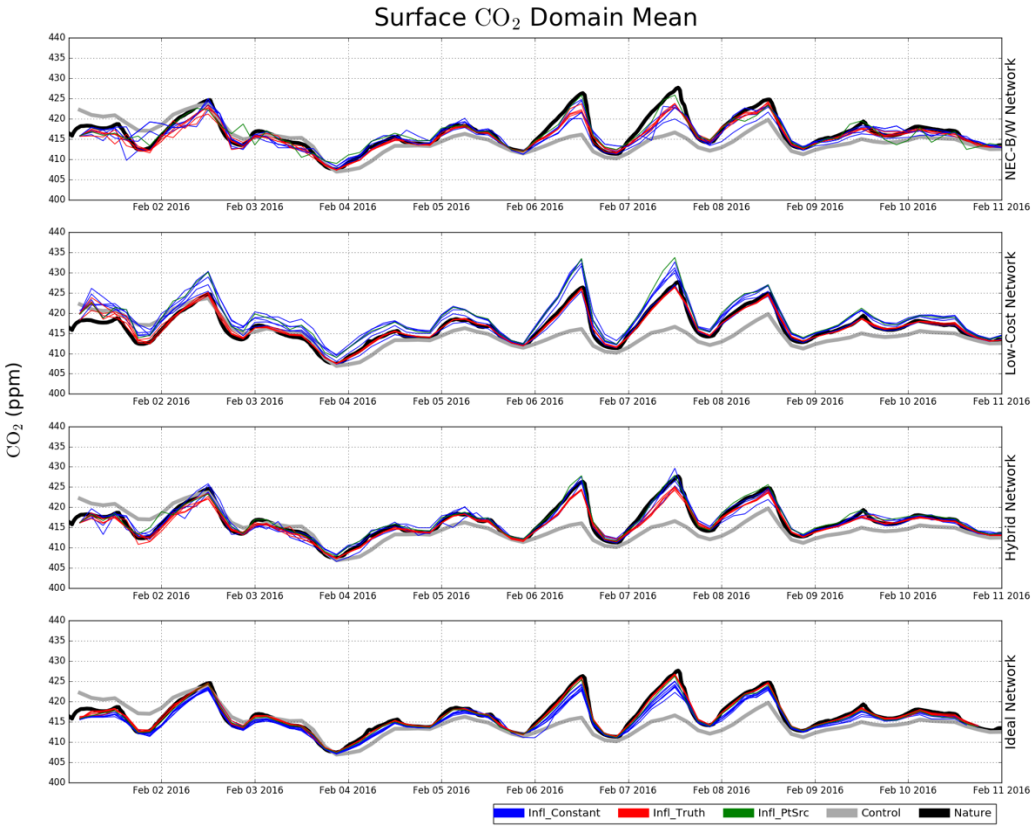

Figure 4.25. Time series of domain averaged surface CO_2 mole fractions (ppm) for the nature run (black), a control simulation without data assimilation (gray), and multiple OSSEs using a variety of additive inflation techniques (Infl_Constant: blue; Infl_Truth: red; Infl_PtSrc: green) and observation localization radii.

4.5.6 Uncertainty

Because LETKF is an ensemble data assimilation method, not only is an analysis generated at each cycle from the mean of the ensemble, but the spread of the ensemble gives an uncertainty estimate of the analysis. Figure 4.26 shows the domain averaged ensemble spread over days 4-8 in the bars (with the error bars showing the standard deviation of the spread over this period) for Infl_Constant on the left, and Infl_Truth on the right, for all four observing networks and the different localization radii. Regardless of inflation method or observing network, there is a clear trend in that as the localization radius increases, the ensemble spread will decrease. This is

likely related to the increase in the number of observations assimilated to better constrain the estimate. With the exception of the 1km radius experiments using the Ideal network, and the three 50 km Infl_Constant experiments for the sparser networks, the ensemble spread averaged across the domain is generally on the order of 5% of the domain total flux value (approximately 350 mol km^{-2} hr^{-1}). However, the ensemble spread for these outlying experiments can be as high as 1400 mol km^{-2} hr^{-1}, a value that is approximately 20% of the true domain averaged flux. While the experiments using Infl_Truth overall have smaller ensemble spread relative to those using Infl_Constant, the differences are not more than 1-2% of the true domain averaged value.

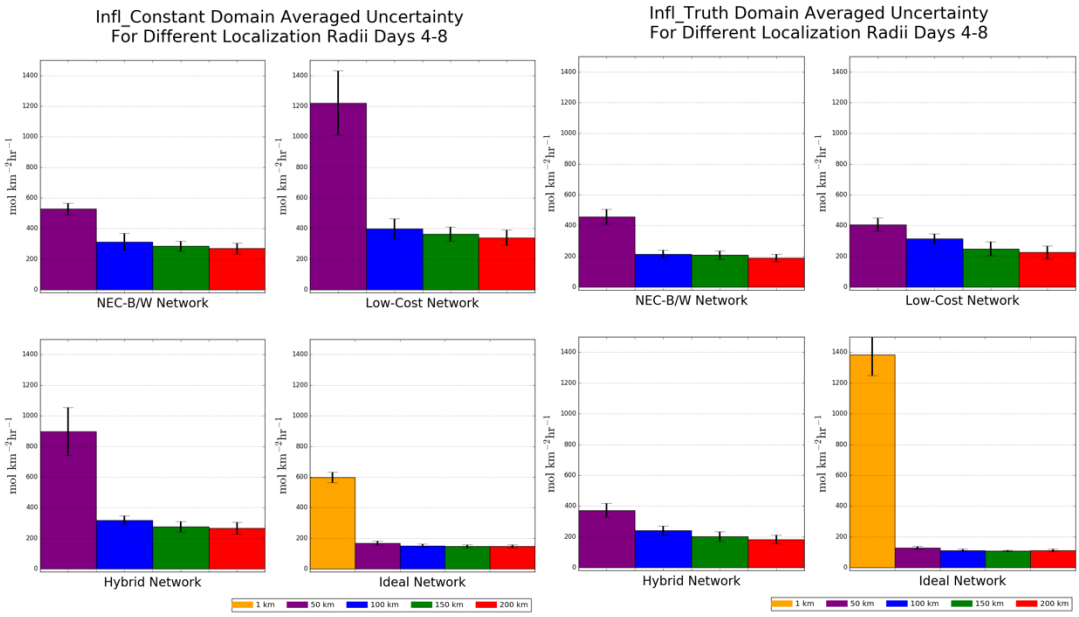

Figure 4.26. Domain averaged ensemble spread of the estimated surface CO_2 flux over days 4-8 shown as the bars (with the error bars showing the standard deviation of the spread over this period) for Infl_Constant on the left, and Infl_Truth on the right, for all four observing networks and the different localization radii.

4.6 Summary and Conclusions

A series of OSSEs are conducted and evaluated to determine not only if using LETKF with WRF-Chem can accurately estimate high-resolution surface CO_2 fluxes in an urban region, but how changes in the ensemble data assimilation system affect the results. Averaged over five days, for an idealized case with a high-accuracy observation at each model grid point, the LETKF system can estimate the domain averaged surface CO_2 flux to within 0.1% of the assumed true value, however this observing network is not feasible due to cost or other implementation limitations. Using three realistic observing network scenarios (20 high-accuracy locations, 200 medium-accuracy locations, or a hybrid network combining both for 220 observing points), it is not possible to get a domain total surface flux estimate averaged over five days to be within even 10% of the true value. When focusing just on the urban center of the modeling domain however, certain configurations can get this five-day average within 10% or even within 5% of the true value. For both the total domain and the urban center, while the temporal mean of the areal sum may not be well resolved, each observing network has an ensemble configuration that can accurately estimate the surface flux for one or more analysis cycles. In fact, for observing networks with Obs_HA observations, the best analysis cycle can be within 1% and within 4% when using 200 Obs_LC observations. This further emphasizes the impact that meteorology plays in the ability to accurately estimate surface fluxes (as described in Chapter 3).

Ensemble data assimilation techniques such as LETKF get their forecast (background) error statistics from the spread of the ensemble members. Thus, the

inflation technique used after each analysis cycle has a significant impact on the quality of the estimated surface fluxes and atmospheric CO_2. The observation localization radius, or the distance from a grid point that an observation can influence the analysis, also plays a significant role in the quality of the estimated surface flux relative to the true values, but this effect is secondary to the inflation technique used. Additionally, while a certain inflation technique tends to work best for all observation network configurations, the localization radius will need to be optimized depending on the quantity and distribution of observations used. The results from these OSSEs suggest that LETKF coupled with an Eulerian meteorological transport model could be a reasonably useful tool for urban greenhouse gas flux estimation, but further evaluation is necessary.

Acknowledgements

I would like to thank Takemasa Miyoshi for making available his LETKF data assimilation code (available at: https://github.com/takemasa-miyoshi/letkf) which was modified for use in this chapter.

Chapter 5 Conclusions and Future Work

5.1 Conclusions Addressing Main Research Questions

The individual chapters in this dissertation examined the viability of using a relatively dense network of low-cost and medium-accuracy CO_2 sensors, coupled with an atmospheric transport model, and an advanced data assimilation system, to estimate anthropogenic CO_2 fluxes in an urban area. In Chapter 2, the performance of a particular low-cost NDIR CO_2 sensor was evaluated, and with corrections for atmospheric variables such as air pressure and temperature, the sensor can measure CO_2 within 2 ppm (at the 95% confidence interval) compared to a reference analyzer when averaged over the entire month. Chapter 3 compared a deterministic simulation of WRF-Chem predicted atmospheric CO_2 to observations. It was argued that while WRF-Chem combined with prior anthropogenic emissions inventories could reasonably reproduce the CO_2 mole fractions averaged over the entire month, there is substantial variation in the quality of the modeled CO_2 mole fractions for any given hour or observing location. Finally, in Chapter 4 a series of OSSEs are performed to determine if using WRF-Chem with an ensemble data assimilation technique could accurately estimate urban CO_2 fluxes with 200 low-cost but medium-accuracy sensors, as well as for other observing networks.

Chapter one featured a number of research questions to be addressed in this dissertation and detailed conclusions for each of these questions are presented in the subsequent subsections.

5.1.1 Are any of the current commercially available and low-cost observing platforms for CO_2 able to resolve the ambient variability both in time and space with sufficient accuracy and precision for use in urban environments?

Chapter 2 features the evaluation of one particular low-cost CO_2 observing platform, a NDIR sensor called the K30 from a Swedish company, SenseAir. The manufacturer's stated accuracy is ±30 ppm ±3 % of the reading, insufficient for ambient CO_2 monitoring as even in urban areas, enhancements are generally on the same order of tens of ppm. However, after co-locating six of these sensors with a high-accuracy gas analyzer, it was found that most of this manufacturer's stated uncertainty was due to an individual offset for each sensor (Fig. 2.4). Just by correcting for this offset, the accuracy of this sensor increases to approximately 1% of the observed value (4-5 ppm). To further reduce the measurement error of these sensors, a multivariate regression analysis was performed including atmospheric pressure, temperature, and water vapor mixing ratio in addition to the assumed true value of CO_2 from the gas analyzer. By including all of these variables in a correction, the measurement error can be reduced to under 2 ppm for 1-minute data (Fig. 2.8). When increasing the averaging window to one hour (commonly done for atmospheric observations of CO_2), the root mean square error can be reduced to as low as 1.5 ppm. Based on the findings from this study, with environmental correction, the K30 sensor can measure variability in atmospheric CO_2 with sufficient precision for urban greenhouse gas monitoring.

5.1.2 Are high-resolution forward modeled simulations of atmospheric CO_2 mole fractions able to accurately resolve the mean and temporal variability found in observed time series?

Overall, WRF-Chem generated mole fraction time series have similar magnitudes and variabilities to the observed mole fractions from the four observation locations used in Chapter 3, and have realistic variations in the synoptic and diurnal cycles (Fig. 3.4). The average bias of WRF-Chem over the month of February while using an ensemble of anthropogenic emissions inventories varies from approximately -3 to 1 ppm across all of the observational data used (Fig. 3.6). However, Section 3.3.2 illustrates that depending on the synoptic meteorological regime in the area, the model performance for CO_2 mole fraction simulation can vary significantly. At any particular site, the differences between the ensemble of simulated CO_2 values and the observed CO_2 can vary from near zero to as high as 100 ppm (Fig. 3.4). When averaged over the entire month, all simulated CO_2 mole fractions are within 8 ppm of each other (Fig. 3.6) representing a range of approximately 2% of the total mole fraction.

5.1.3 Is the error associated with differences in emissions inventories greater than the error from meteorological transport and dispersion when comparing simulated CO_2 to observations?

In addition to the synoptic meteorology, variations in emissions inventories are also reflected in the predicted CO2 mole fractions as shown in the mean biases described in Fig. 3.6. For the outermost domain (d01), the areal sum of the hourly averaged emissions is similar in magnitude (all within 10 % of the mean of the 4 inventories) (Fig. 3.3; Table 3.3). However, in the innermost domain (d03), the areal sum of FFDAS is over 36 % higher than the mean and the lowest inventory EDGAR

is 20% below the mean of the inventories. Even though each inventory has a similar national or global total, there can be substantial differences between them for any given region due to the emission disaggregation methods used (e.g., Hutchins et al. 2016; Oda et al. 2018), a problem being studied further in other works (e.g., Fischer et al., 2017). The differences in totals and local sources could also be attributed to differing methodologies and datasets included in each emissions inventory, including the exact location of point sources and grid cell locations, among other things. These differences are generally reflected in the simulated CO_2 levels, with FFDAS being the highest averaged over the entire month, and EDGAR the lowest.

The mean absolute error of the simulated CO_2 is roughly a factor of two larger than the mean standard deviation at the urban sites, with the ratio of the two ranging from 1.64 to 2.58 for all hours, and 1.57 to 1.89 for afternoon hours only, depending on the site and inlet level. This result suggests that on average, factors common to all five tracers (meteorological error, background error, or error in the biosphere tracer) contribute more to the overall model performance than the choice of anthropogenic emissions inventory. Given the low bias (~2 ppm) at SNP, the extent of our largest WRF domain, and the small relative contribution of the VEGAS tracer to the monthly mean values, we expect that the errors shown in Fig. 3.10 are dominated by meteorological conditions during winter. This conclusion is further supported by the two examples illustrated in Section 3.3.2. However, it is important to note that these meteorological conditions or errors can exacerbate the differences in the emissions inventories as well, like shown in Fig. 3.7.

5.1.4 Can an ensemble of forward tracer model simulations be used with in situ observations of atmospheric CO_2 and data assimilation techniques to estimate high resolution surface fluxes of carbon in urban areas?

Chapter 4 featured a series of OSSEs using WRF-Chem coupled with LETKF to demonstrate the possibility of using ensemble data assimilation with Eulerian transport and dispersion models to generate 3D analysis fields of atmospheric CO_2 as well as to estimate surface carbon fluxes. Using an idealized case of a high accuracy observation at every model grid point and an additive inflation method that includes the spatial distribution and relative magnitudes of the assumed true fluxes, the data assimilation system can estimate the domain averaged surface CO_2 flux to within 0.5% of the assumed correct value (Fig. 4.11). This holds true whether looking at the domain total or just the urban center area of interest (Fig. 4.18). However, Sect. 4.5.3 and Fig. 4.22 show that while the other observing networks can not accurately estimate the surface flux over a five-day average, there are optimum analysis cycles where the error can be within 1% for a network featuring high-accuracy observations to 4 % for the network with 200 low-cost observations.

5.1.5 How does the estimation of surface fluxes using a dense network of lower accuracy observations compare to using a sparse network of high accuracy observations?

In Chapter 4, four separate hypothetical observation networks of in situ CO_2 concentrations are used: one with a sparse network of high accuracy observations, one with a dense network of lower accuracy observations, a network combining the two previous networks, and an idealized network with a high accuracy observation at each model grid point. When using an ensemble with no spatial information introduced into the forecast error covariances (Table 4.3), the network of 200 lower

accuracy observations significantly out performs the 20 high accuracy observation network. While the high accuracy network only estimates a domain averaged flux that is approximately 30% of the true value, when using the 200 lower accuracy observations the system can get the value to approximately 90% of the true value (and the true value is within the range of uncertainty for this experiment). When introducing prior spatial information into the system (Table 4.4), the differences are much less dramatic (the difference between the two is less than 5%), however the dense, but lower accuracy network still out performs the sparse, high accuracy observation network. These results are based on a five-day average of the surface flux, when considering just the analysis cycle with the mean value closest to the truth (Sect. 4.5.3) there is a distinct advantage to using the high-accuracy observations. The 20 high accuracy observation network out performs the 200 lower accuracy sensors, with a lowest mean error of within 1% vs 4%.

5.1.6 Is there any added benefit to the surface flux estimation by creating a hybrid network containing both the high accuracy and low accuracy observations over just using the sparse high accuracy observation network?

When using the additive inflation method without any spatial information (Sect. 4.4.2.1), this combined/hybrid network performs better than the sparse, but high accuracy network, but worse than the dense network. Perhaps unsurprisingly, the results are somewhere in between these other two networks for most of the experiments performed in Chapter 4. In almost all cases, the hybrid network underestimates the domain averaged CO_2 flux compared to the true value, and the lower accuracy network's estimation. This is likely due to the fact that the high accuracy observations generally are located in more rural areas, and at higher heights

above ground level, so the concentrations observed at these sites will tend to be lower than that of those from the near-surface lower-accuracy observations limited to the urban areas. When looking at each experiment's best performing analysis period (Sect. 4.5.3; Fig. 4.22), we can see a slight decrease in the mean absolute error of the analysis (a decrease of approximately 0.1% of the mean value) when adding 200 low accuracy sensors to the 20 high accuracy sensor network, but this decrease in error is not much for this particular cycle. It is clear that overall, the performance of this data assimilation system is improved across all different configurations when including this additional network of low-cost and lower accuracy sensors to complement a high accuracy network, but further evaluation needs to be done to determine the extent of this improvement.

5.1.7 How do changes in the configuration of the data assimilation system affect the surface flux estimates?

For a given observation network, the estimate of surface carbon flux can change significantly depending on the specific data assimilation configurations chosen. Section 4.4 evaluates the estimated surface flux while varying the additive inflation method used for the ensemble after each analysis cycle, the effect changing the observation localization radius has, the amount/size of the inflation at each cycle, as well as changing the amount and distribution of observations. When varying the inflation technique for a fixed localization radius (Table 4.2), the differences between the three techniques can be as small as 10% or as large as a factor of 5 differences between the smallest and the largest of the three techniques. The spatial correlations

between the same spatial pixel for the true fluxes and the analyzed fluxes are also much better when using Infl_Truth compared to Infl_Constant (Figs. 4.20, 4.21). For a fixed inflation technique (Tables 4.3, 4.4), the changes across different observation localization radii are not as dramatic but still significant. The difference between an accurate estimation of the surface CO_2 flux and an underestimation can be solely because of the choice of localization radius (Figs. 4.11, 4.18). Additionally, the amount of the inflation perturbations to the ensemble members (Fig. 4.13), the exact location of observations (Fig. 4.14), and a relatively small increase in the number of observations (Fig. 4.15), each affect the results by a non-trivial amount but they are secondary responses compared to that from changing the inflation techniques and localization radii.

5.2 Future Research Directions

The research conducted as part of this dissertation is largely interdisciplinary, as it incorporates a significant component of instrumentation evaluation, mesoscale modeling of meteorology and trace gases, and ensemble data assimilation. As such, to describe the future research that this work should inspire, it is perhaps best to separate these directions into separate sections. Section 5.2.1 suggests ways to improve low-cost CO_2 observing platforms for use in urban monitoring. Section 5.2.2 describes additional work that needs to be investigated for regional and urban CO_2 modeling. Finally, Section 5.2.3 identifies future tests and experiments that should be conducted to further demonstrate the viability of ensemble data assimilation as a tool for high resolution CO_2 flux estimation.

5.2.1 Future Direction for Low-Cost Sensors

Chapter 2 concluded that the K30 low-cost NDIR CO_2 sensor could have an accuracy of better than 2 ppm when compared to a reference analyzer after correcting for environmental factors. However, this analysis was only over a 30-day period and did not evaluate if the sensor features any long-term drift. Thus, a long-term evaluation of the K30 sensor, for at least six months but a year is preferred, is needed where the sensor is co-located with a reference analyzer to determine if there is long-term drift in the sensor, and if the calibration coefficients computed during the first few weeks are still valid throughout the evaluation period. This could provide insight on the stability of the instrument and how often a recalibration may be required for adequate results. Given the relative small size and power consumption requirements of this sensor, future work could also entail evaluating its use in other observing platforms besides surface observations, including placing them on unmanned aerial vehicles or in a sensor package for vertical profiles using a balloon (radiosonde).

Additionally, if a method could be devised to regularly calibrate these sensors at relatively low-cost with reference gas standards, either one for baseline drift, or preferably two different values for a two-point calibration, this could further improve both the accuracy of the observations, but also the legitimacy of any datasets that include the observations. Longer term, new technology will almost certainly be developed that can either improve on existing observing methods such as NDIR, or perhaps a new method, that can either decrease the cost and size of an instrument or

increase the accuracy and precision. Efforts are already underway at SenseAir to create a new NDIR CO_2 observing platform that may be comparable to the K30 in size and cost but feature improved performance. At the very least, methods to control the influence environmental factors have on the measured CO_2 could have a significant impact on the real-world performance of low-cost observing platforms.

5.2.2 Future Direction for Urban CO_2 Modeling

Numerical weather prediction is a constantly evolving field, with improvements in model physics and dynamics coming every year, and as computers become faster, the feasible resolution of simulations increases as well. Related to the issue at hand of urban CO_2 modeling, there are numerous areas of improvement in the short term that can be addressed. First, implementing advanced data assimilation techniques as well as just including new observing platforms of meteorological variables into the meteorological analysis can improve the prediction of urban CO_2 mole fractions. Analysis nudging was not used in Chapter 3 because it is believed to not fully conserve mass, and thus ways to improve the meteorology without affecting the CO_2 tracers are needed. If the modeling community can improve the transport and dispersion within the model, then the simulated CO_2 will almost certainly more realistically represent the true CO_2 field. Secondly, while Chapter 3 found this to be secondary to the model meteorology, there is still uncertainty and variability in anthropogenic emissions inventories. As new technologies emerge, and areas shift from reliance on coal to natural gas, or from fossil fuels to renewable energy, or if a developing country builds new emissions sources, then there will still be a need for an

accurate and up to date emissions inventory to use as input to the model. Most of these inventories are at too low resolution, either spatially or temporally, to accurately represent the variability observed in urban areas, so increasing the resolution would be a welcome improvement.

The model simulation in Chapter 3 was for February 2016. This month was chosen not only because of data availability, but because the biospheric CO_2 signal was assumed to be small compared to the anthropogenic emissions. Because of this, there was no evaluation of the VEGAS biospheric flux model, which would be key for accurate CO_2 mole fraction simulation during the growing season. Thus, one obvious first next step would be to recreate the analysis from Chapter 3 but with a summer month rather than a winter month. Not only would this help to evaluate the modeling framework with the biospheric flux added, but it would also show how the results differ (or are similar) due to the summer meteorology (more convection, fewer mid-latitude storms).

Continuing with the differences in meteorology between summer and winter in the mid-latitudes, Chapter 3 suggests that the overarching synoptic meteorology setup may play a role in the ability for the model to accurately predict CO_2 mole fractions. Some days the modeled CO_2 error is quite small (under 1%), but for others the error can be well over 10% of the observed CO_2 mole fraction. Future work should investigate this further and see if it applies to multiple regions and over all seasons/months. Hopefully, this future analysis could lead to criteria or a technique for determining which periods the model can accurate represent the CO_2 mole

fractions, both for high-resolution modeling, but also for use in surface flux estimation.

5.2.3 Future Direction for Ensemble DA CO_2 Flux Estimation

Chapter 4 performed a number of OSSEs to determine the validity of using LETKF with WRF-Chem and in situ observations of CO_2 mole fractions to estimate surface CO_2 fluxes. One of the main findings of this chapter was that the configuration of the data assimilation system, particularly the inflation technique, has a significant impact on the quality of the CO_2 flux analysis. The forecast or background error is very important for data assimilation systems to accurately generate a new analysis. First, to better understand the uncertainty of the estimations, one can look more closely at the ensemble spread in the experiments presented in Chapter 4 as well as in subsequent studies, as the LETKF code provides the analysis mean and spread as output at each cycle, this is briefly presented in Chapter 4.5.6. Further OSSEs should be performed using this data assimilation system to evaluate the results with additional inflation techniques, both the additive inflation methods described in Chapter 4, but also with new ones such as multiplicative inflation or relaxation to prior. In addition to the inflation techniques, further experiments should be conducted to evaluate additional observing networks to find a configuration (or multiple ones) that can best estimate the surface CO_2 fluxes. Then with these new observing network configurations and inflation techniques, greater care should be taken to find the optimum localization radius (do a sensitivity test at 5 km intervals rather than 50 km, for example) for the surface flux estimation.

As with the high-resolution modeling of CO_2, the flux estimation OSSEs should also be conducted in a summer period as well as a winter period, to see how the biospheric fluxes can affect the results. With this, there will also need to be an error associated with the biospheric flux, and care will have to be taken to separate out the separate signals of the fluxes from the observations which are just the total CO_2 mole fraction. Finally, after these additional OSSEs are performed, the best performing configuration(s) should be used with real observations of in situ CO_2 mole fractions to estimate the surface CO_2 flux, with the information gathered from the OSSEs used to generate the uncertainty associated with those estimates. LETKF coupled with WRF-Chem has shown to be a viable option for high-resolution CO_2 flux estimation and by using it along with traditional inverse-based flux estimation methods, urban carbon cycle scientists will be able to better estimate both the true surface flux, but also the uncertainty associated with these estimates.

5.3 Final Thoughts

Currently, the United States Environmental Protection Agency uses continuous emissions monitoring systems (CEMS) to estimate the emissions of trace gases, including CO_2, from smokestacks. The CEMS systems are calibrated yearly with a relative accuracy test audit, but this only as the name suggests, provides relative accuracy calibrations. Evaluations conducted by NIST suggest that the uncertainty in emissions estimates from CEMS can be as large as 10%, with efforts underway to reduce this uncertainty to 1% by developing new techniques to measure the turbulent flow inside of the smokestacks (Johnson et al., 2015). The results presented in Chapter 4 suggest that using LETKF with WRF-Chem and various

observing networks can also estimate surface CO_2 flux to within 10% of the true value.

As less-wealthy countries continue to develop their economies, and with that increase their consumption of fossil fuels, there will be a need to estimate their greenhouse gas emissions. In these countries, such as those in Africa or the Indian subcontinent, the local governments may not be able to dedicate significant resources to climate change mitigation, particularly relative to places like Europe, China, or the United States. In these developing nations are where emissions estimations using low-cost sensor networks may be most useful. The OSSEs presented in Chapter 4 show that flux estimates using 200 low-cost sensors provide comparable results to that of using 20 high-accuracy instruments, for a total installation cost comparable to that of just one of the high-accuracy locations. Thus, while some tradeoffs in accuracy may be made, an estimation with a slightly larger uncertainty is going to be much more useful than no estimation at all because of prohibitive cost.

Improving our estimates of CO_2 emissions will be a crucial area of research with implications for the future of Earth's climate, geopolitical conflicts, and the survival of mankind. While the current uncertainties of emissions estimates are still on the order of 10%, with additional observing systems and new estimation techniques, including those presented in this work, future estimations will almost certainly be closer to the true emissions quantity. Accurate CO_2 flux estimation is key from a political standpoint; as most countries agree to mitigate emissions, accountability of the pledged reduction in emissions over time requires accurate estimates to confirm or dispute each nation's progress. Additionally, in order to better

estimate the future global climate, an accurate estimation of global CO_2 emissions is key to understanding which of the RCP scenarios we are most likely to achieve. Anthropogenic climate change is settled science, however what is far from settled is the magnitude that our planet's climate will change in the coming decades. While there are many factors that will contribute to the net increase in global averaged temperatures, including aerosols and clouds, better estimation of CO_2 emissions is a key step towards understanding the future atmospheric composition and its impact on our climate.

Bibliography

Allan, D. W.: Statistics of Atomic Frequency Standards, Proceedings of the Institute of Electrical and Electronics Engineers, 54, 221, doi: 10.1109/proc.1966.4634, 1966.

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, Atmos. Meas. Tech., 7, 647-687, 10.5194/amt-7-647-2014, 2014.

Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., Huang, J., Elvidge, C., and Baugh, K.: A multiyear, global gridded fossil fuel CO2 emission data product: Evaluation and analysis of results, Journal of Geophysical Research: Atmospheres, 119, 10,213-210,231, 10.1002/2013JD021296, 2014.

Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO_2, Tellus B, 58, 359-365, doi:10.1111/j.1600-0889.2006.00218.x, 2006.

Beck, V., T. Koch, R. Kretschmer, J. Marshall, R. Ahmadov, C. Gerbig, D. Pillai, and M. Heimann, 2011: The WRF Greenhouse Gas Model (WRF-GHG) Technical Report No. 25, Max Planck Institute for Biogeochemistry, Jena, Germany, available online at http://www.bgc-jena.mpg.de/bgc-systems/index.shtml

Bosch Sensortec: BME280 Digital Pressure Sensor Datasheet, available at: https://cdn-shop.adafruit.com/datasheets/BST-BME280_DS001-10.pdf (last access: 7 June 2016), 2015.

Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., and Monfray, P.: Inverse modeling of annual atmospheric CO_2 sources and sinks: 1. Method and control inversion, Journal of Geophysical Research: Atmospheres, 104, 26161-26178, doi:10.1029/1999JD900342, 1999.

Breón, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., Dieudonne, E., Lopez, M., Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements, Atmospheric Chemistry and Physics, 15, 1707-1724, 10.5194/acp-15-1707-2015, 2015.

Briber, B., Hutyra, L., Dunn, A., Raciti, S., and Munger, J.: Variations in Atmospheric CO_2 Mixing Ratios across a Boston, MA Urban to Rural Gradient, Land, 2, 304, 2013.

Carbontracker Team: Compilation of near real time atmospheric carbon dioxide data provided by NOAA and EC; obspack_co2_1_NRT_v3.3_2017-04-19; NOAA Earth

System Research Laboratory, Global Monitoring Division. http://doi.org/10.15138/G3G01J, 2017.

Chatterjee, A., Michalak, A. M., Anderson, J. L., Mueller, K. L., and Yadav, V.: Toward reliable ensemble Kalman filter estimates of CO_2 fluxes, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2012JD018176, 2012.

Chatterjee, A., Engelen, R. J., Kawa, S. R., Sweeney, C., and Michalak, A. M.: Background error covariance estimation for atmospheric CO_2 data assimilation, Journal of Geophysical Research: Atmospheres, 118, 10,140-110,154, doi:10.1002/jgrd.50654, 2013.

Chatterjee, A. and Michalak, A. M.: Technical Note: Comparison of ensemble Kalman filter and variational approaches for CO_2 data assimilation, Atmos. Chem. Phys., 13, 11643-11660, https://doi.org/10.5194/acp-13-11643-2013, 2013.

Chou, M.-D., Suarez, M. J., Liang, X.-Z., Yan, M. M.-H., and Cote, C.: A thermal infrared radiation parameterization for atmospheric studies, 2001.

Ciais P, Rayner P, Chevallier F, Bousquet P, Logan M, Peylin P, Ramonet M (2010) Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives. Clim Change 103(1-2):69-92. https://doi.org/10.1007/s10584-010-9909-3

Clarke L., K. Jiang, K. Akimoto, M. Babiker, G. Blanford, K. Fisher-Vanden, J.-C. Hourcade, V. Krey, E. Kriegler, A. Löschel, D. McCollum, S. Paltsev, S. Rose, P.R. Shukla, M. Tavoni, B.C.C. van der Zwaan, and D.P. van Vuuren: Assessing Transformation Pathways. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

Deng, A., T. Lauvaux, K. J. Davis, B. J. Gaudet, N. Miles, S. J. Richardson, K. Wu, D. P. Sarmiento, R. M. Hardesty, and T. A. Bonin (2017), Toward reduced transport errors in a high resolution urban CO_2 inversion system, Elem Sci Anth, 5.

Eugster, W., and Kling, G. W.: Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies, Atmospheric Measurement Techniques, 5, 1925-1934, 10.5194/amt-5-1925-2012, 2012.

Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A. J., Diaz-Isaac, L. I., Duren, R. M., Fischer, M. L., Gerbig, C., Gurney, K. R., Huang, J. H., Jeong, S., Li, Z. J., Miller, C. E., O'Keeffe, D., Patarasuk, R., Sander, S. P., Song, Y., Wong, K. W., and Yung, Y. L.: Los Angeles megacity: a high-resolution land-atmosphere modelling system for urban CO_2 emissions, Atmospheric Chemistry and Physics, 16, 9019-9045, 10.5194/acp-16-9019-2016, 2016.

Fischer, M. L., Parazoo, N., Brophy, K., Cui, X., Jeong, S., Liu, J., Keeling, R., Taylor, T. E., Gurney, K., Oda, T., and Graven, H.: Simulating estimation of California fossil fuel and biosphere carbon dioxide exchanges combining in situ

tower and satellite column observations, Journal of Geophysical Research: Atmospheres, 122, 3653-3671, 10.1002/2016JD025617, 2017.

Gas Sensing Solutions: COZIR Ultra Low Power Carbon Dioxide Sensor, available at: http://www.gassensing.co.uk/media/1050/cozir_ambient_datasheet_gss.pdf (last access: 29 December 2015), 2014.

Gately, C. K., & Hutyra, L. R. (2017). Large uncertainties in urban-scale carbon emissions. Journal of Geophysical Research: Atmospheres, 122, 11,242–11,260. https://doi.org/10.1002/2017JD027359

Gemmill, W., Katz, B., and Li, X.: Daily real-time global sea surface temperature-high resolution analysis at NOAA/NCEP, NCEP Off. Note, 260, 39, 2007.

General Electric: Telaire T6615 Sensor Dual Channel Module, available at: http://www.avnet-abacus.eu/fileadmin/user_upload/Products_Menu/Amphenol/AmphenolAdvancedSensors_CO2_double_channel_module.pdf (last access: 29 December 2015), 2011.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, Journal of Geophysical Research: Atmospheres, 108, doi:10.1029/2003JD003770, 2003.

Gourdji, S. M., Mueller, K. L., Schaefer, K., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, Journal of Geophysical Research: Atmospheres, 113, doi:10.1029/2007JD009733, 2008.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957-6975, 10.1016/j.atmosenv.2005.04.027, 2005.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de la Rue du Can, S.: High Resolution Fossil Fuel Combustion CO_2 Emission Fluxes for the United States, Environmental Science & Technology, 43, 5535-5541, 10.1021/es900806c, 2009.

Gurney, K.R., Razlivanov, I., Song, Y. Zhou, Y., Benes, B., Abdul-Massih, M.: Quantification of fossil fuel CO_2 at the building/street scale for a large US city, Environmental Science and Technology, 46, 12194-12202, dx.doi.org/10.1021/es3011282, 2012.

Holstius, D. M., Pillarisetti, A., Smith, K. R., and Seto, E.: Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California, Atmospheric Measurement Techniques, 7, 1121-1131, 10.5194/amt-7-1121-2014, 2014.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, Monthly Weather Review, 132, 103-120, 10.1175/1520-0493(2004)132<0103:aratim>2.0.co;2, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, Monthly Weather Review, 134, 2318-2341, 10.1175/mwr3199.1, 2006.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D-Nonlinear Phenomena, 230, 112-126, 10.1016/j.physd.2006.11.008, 2007.

Hurst, S., Durant, A. J., and Jones, R. L.: A low cost, disposable instrument for vertical profile measurements of atmospheric CO_2, Chemistry Research Project Report, Centre for Atmospheric Science, Department of Chemistry, University of Cambridge, 2011.

Hutchins, M. G., Colby, J. D., Marland, G., and Marland, E.: A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States, Mitigation and Adaptation Strategies for Global Change, 1-26, 10.1007/s11027-016-9709-9, 2016.

IEA, 2012. CO_2 emissions from fuel combustion 1971–2010, 2012 Edition, Paris: International Energy Agency (IEA), 2012.

Intergovernmental Panel on Climate Change (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, In T. F. Stocker, et al. (Eds.), (1535 pp.). Cambridge, U. K., and New York: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

Johnson, Aaron N., Rodney A. Bryant, Tamae Maeda Wong, James R. Whetstone, Toralf Dietz, Donald Giel, David Elam Jr., Scott Swiggard, Woong Kang, Iosif Shinder, Eric Harman, Hsin-Hung Lee, Liang Zhang, Keith Gillis, and Richard Grot: Measurement Challenges and Metrology for Monitoring CO_2 Emissions from Smokestacks – Workshop Summary, NIST Special Publication 1201, http://dx.doi.org/10.6028/NIST.SP.1201, 2015.

Jones E, Oliphant E, Peterson P, et al.: SciPy: Open Source Scientific Tools for Python, http://www.scipy.org/ (last access: 10 October 2016), 2001.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170-181, 10.1175/1520-0450(2004)043<0170:tkcpau>2.0.co;2, 2004.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, Tellus A: Dynamic Meteorology and Oceanography, 59, 758-773, 10.1111/j.1600-0870.2007.00261.x, 2007.

Kang, J. S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K.: "Variable localization" in an ensemble Kalman filter: Application to the carbon cycle data assimilation, Journal of Geophysical Research: Atmospheres, 116, doi:10.1029/2010JD014673, 2011.

Kang, J. S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2012JD018259, 2012.

Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., and Meijer, H. A.: Atmospheric CO_2 and (13)CO_2 exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications, History of Atmospheric CO_2 and Its Effects on Plants, Animals, and Ecosystems, 177, 83-113, 2005.

Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO_2 emissions: Minimum site location requirements for the Los Angeles megacity, Journal of Geophysical Research-Atmospheres, 118, 1-8, 10.1002/jgrd.50135, 2013.

Kunz, M., Lavric, J. V., Gerbig, C., Tans, P., Neff, D., Hummelgård, C., Martin, H., Rödjegård, H., Wrenger, B., and Heimann, M.: COCAP: a carbon dioxide analyser for small unmanned aircraft systems, Atmos. Meas. Tech., 11, 1833-1849, 10.5194/amt-11-1833-2018, 2018.

Langridge, J. M., Ball, S. M., Shillings, A. J. L., and Jones, R. L.: A broadband absorption spectrometer using light emitting diodes for ultrasensitive, in situ trace gas detection, Review of Scientific Instruments, 79, 14, doi: 10.1063/1.3046282, 2008.

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., and O'Keefe, D.: High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), Journal of Geophysical Research: Atmospheres, 2016.

Lee, T. R., De Wekker, S. F. J., Andrews, A. E., Kofler, J., and Williams, J.: Carbon dioxide variability during cold front passages and fair weather days at a forested mountaintop site, Atmospheric Environment, 46, 405-416, https://doi.org/10.1016/j.atmosenv.2011.09.068, 2012.

Leip, A., Skiba, U., Vermeulen, A., and Thompson, R. L.: A complete rethink is needed on how greenhouse gas emissions are quantified for national reporting, Atmospheric Environment, 174, 237-240, https://doi.org/10.1016/j.atmosenv.2017.12.006, 2018.

Lewis, Alastair, W. Richard Peltier, and Erika von Schneidemesser: Low-cost sensors for the measurement of atmospheric composition: Overview of topic and future applications, World Meteorological Organization Technical Report, https://www.wmo.int/pages/prog/arep/gaw/documents/Low_cost_sensors_post_review_final.pdf, 2018.

Liu, J., Bowman, K. W., and Lee, M.: Comparison between the Local Ensemble Transform Kalman Filter (LETKF) and 4D-Var in atmospheric CO_2 flux inversion with the Goddard Earth Observing System-Chem model and the observation impact diagnostics from the LETKF, Journal of Geophysical Research: Atmospheres, 121, 13,066-013,087, doi:10.1002/2016JD025100, 2016.

Lopez-Coto, I., Ghosh, S., Prasad, K., and Whetstone, J.: Tower-based greenhouse gas measurement network design—The National Institute of Standards and Technology North East Corridor Testbed, Advances in Atmospheric Sciences, 34, 1095-1105, 10.1007/s00376-017-6094-6, 2017.

Los Gatos Research: Fast Greenhouse Gas Analyzer (Enhanced Performance Model) Datasheet, available at: http://www.lgrinc.com/documents/LGR_FGGA_Datasheet.pdf (last access: 29 December 2015), 2013.

Martin, C. R., Zeng, N., Karion, A., Dickerson, R. R., Ren, X., Turpie, B. N., and Weber, K. J.: Evaluation and environmental correction of ambient CO_2 measurements from a low-cost NDIR sensor, Atmospheric Measurement Techniques, 10, 2383-2395, 10.5194/amt-10-2383-2017, 2017.

McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, Proceedings of the National Academy of Sciences of the United States of America, 109, 8423-8428, 10.1073/pnas.1116645109, 2012.

McRae, J. E., and Graedel, T. E.: Carbon Dioxide in the Urban Atmosphere: Dependencies and Trends, J. Geophys. Res.-Oceans, 84, 5011-5017, 10.1029/JC084iC08p05011, 1979.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American Regional Reanalysis, Bulletin of the American Meteorological Society, 87, 343-360, 10.1175/bams-87-3-343, 2006.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, Journal of Geophysical Research: Atmospheres, 109, doi:10.1029/2003JD004422, 2004.

Miles, N. L., Richardson, S. J., Lauvaux, T., Davis, K. J., Balashov, N. V., Deng, A., Turnbull, J. C., Sweeney, C., Gurney, K. R., and Patarasuk, R.: Quantification of urban atmospheric boundary layer greenhouse gas dry mole fraction enhancements in the dormant season: Results from the Indianapolis Flux Experiment (INFLUX), Elem Sci Anth, 5, 2017.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16663-16682, 10.1029/97JD00237, 1997.

Mueller, K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, Journal of Geophysical Research: Atmospheres, 113, doi:10.1029/2007JD009734, 2008.

Mueller, K., Yadav, V., Lopez-Coto, I., Karion, A., Gourdji, S., Martin, C., & Whetstone, J.: Siting background towers to characterize incoming air for urban greenhouse gas estimation: A case study in the Washington, DC/Baltimore area. Journal of Geophysical Research: Atmospheres, 123. 10.1002/2017JD027364, 2018.

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., and Deng, F.: Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets, Journal of Geophysical Research: Atmospheres, 118, 917-933, 10.1029/2012JD018196, 2013.

Nehrkorn, T., Henderson, J., Leidner, M., Mountain, M., Eluszkiewicz, J., McKain, K., and Wofsy, S.: WRF simulations of the urban circulation in the Salt Lake City area for CO_2 modeling, Journal of Applied Meteorology and Climatology, 52, 323-340, 2013.

National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory, Global Monitoring Division, Trends in Atmospheric Carbon Dioxide, https://www.esrl.noaa.gov/gmd/ccgg/trends/full.html, Last accessed: January 5, 2018.

Oda, T., and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543-556, 10.5194/acp-11-543-2011, 2011.

Oda, T., and Maksyutov, S. (2015), ODIAC Fossil Fuel CO2 Emissions Dataset (Version name: ODIAC2016), Center for Global Environmental Research, National Institute for Environmental Studies, doi:10.17595/20170411.001. (Reference date : 2017/02/02)

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic Carbon dioxide (CO2), version 2016 (ODIAC2016): A global monthly fossil fuel CO2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth Syst. Sci. Data, 10, 87-107, https://doi.org/10.5194/essd-10-87-2018, 2018

Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N., and Peters, J. A. H. W.: Recent trends in global greenhouse gas emissions:regional trends 1970–2000 and spatial distribution of key sources in 2000, Environmental Sciences, 2, 81-99, 10.1080/15693430500400345, 2005.

Parazoo, N. C., Denning, A. S., Kawa, S. R., Corbin, K. D., Lokupitiya, R. S., and Baker, I. T.: Mechanisms for synoptic variations of atmospheric CO_2 in North America, South America and Europe, Atmos. Chem. Phys., 8, 7239-7254, https://doi.org/10.5194/acp-8-7239-2008, 2008.

Pataki, D. E., Bowling, D. R., and Ehleringer, J. R.: Seasonal cycle of carbon dioxide and its isotopic composition in an urban atmosphere: Anthropogenic and biogenic effects, Journal of Geophysical Research-Atmospheres, 108, 8, doi: 10.1029/2003jd003865, 2003.

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations, Journal of Geophysical Research: Atmospheres, 110, doi:10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Petron, G., and Hirsch, A. I.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proceedings of the National Academy of Sciences, 104, 18925-18930, 2007.

Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., Ciais, P., and AEROCARB contributors: Daily CO_2 flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology, Atmos. Chem. Phys., 5, 3173-3186, https://doi.org/10.5194/acp-5-3173-2005, 2005

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R. P., Lv, Q., Hannigan, M., and Shang, L.: The next generation of low-cost personal air quality sensors for quantitative exposure monitoring, Atmospheric Measurement Techniques, 7, 3325-3336, 10.5194/amt-7-3325-2014, 2014.

Poulida, O., R. R. Dickerson, B. G. Doddridge, J. Z. Holland, R. G. Wardell, and J. G. Watkins: Trace Gas Concentrations and Meteorology in Rural Virginia: 1. Ozone and Carbon-Monoxide, Journal of Geophysical Research-Atmospheres, 96(D12), 22461-22475, 1991.

Raspberry Pi Foundation: Raspberry Pi Hardware Documentation, available at: https://www.raspberrypi.org/documentation/hardware/raspberrypi/ (last access: 29 December 2015), 2015.

Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P., and Koffi, E.: A new global gridded data set of CO_2 emissions from fossil fuel combustion: Methodology and evaluation, Journal of Geophysical Research: Atmospheres, 115, n/a-n/a, 10.1029/2009JD013439, 2010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Dool, H. v. d., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, Journal of Climate, 27, 2185-2208, 10.1175/jcli-d-12-00823.1, 2014.

Sargent, M., Barrera, Y., Nehrkorn, T., Hutyra, L. R., Gately, C. K., Jones, T., McKain, K., Sweeney, C., Hegarty, J., Hardiman, B., and Wofsy, S. C.: Anthropogenic and biogenic CO_2 fluxes in the Boston urban region, Proceedings of the National Academy of Sciences, 10.1073/pnas.1803715115, 2018.

SenseAir : CO_2 Engine K30 Specification, available at: http://www.senseair.com/wp-content/uploads/2015/03/CO2-Engine-K30_PSP110-R7.pdf (last access: 29 December 2015), 2007

Shusterman, A. A., Teige, V., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmos. Chem. Phys. Discuss., 2016, 1-23, 10.5194/acp-2016-530, 2016.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, 2008.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 2004,

Thiébaux, J., Rogers, E., Wang, W., and Katz, B.: A new high-resolution blended real-time global sea surface temperature analysis, Bulletin of the American Meteorological Society, 84, 645-656, 2003.

Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., Davis, K. J., Lauvaux, T., Miles, N. L., Richardson, S. J., Cambaliza, M. O., Shepson, P. B., Gurney, K., Patarasuk, R., and Razlivanov, I.: Toward quantification and source sector identification of fossil fuel CO_2 emissions from an urban area: Results from the INFLUX experiment, Journal of Geophysical Research-Atmospheres, 120, 292-312, 10.1002/2014jd022555, 2015.

Turner, A. J., Shusterman, A. A., McDonald, B. C., Teige, V., Harley, R. A., and Cohen, R. C.: Network design for quantifying urban CO_2 emissions: assessing trade-offs between precision and network density, Atmos. Chem. Phys., 16, 13465-13475, doi: 10.5194/acp-16-13465-2016, 2016.

United Nations, Department of Economic and Social Affairs, Population Division (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352).

van den Bossche, M., Rose, N. T., and De Wekker, S. F. J.: Potential of a low-cost gas sensor for atmospheric methane monitoring, Sensors and Actuators B: Chemical, 238, 501-509, 10.1016/j.snb.2016.07.092, 2017.

Verhulst, K. R., Karion, A., Kim, J., Salameh, P. K., Keeling, R. F., Newman, S., Miller, J., Sloop, C., Pongetti, T., Rao, P., Wong, C., Hopkins, F. M., Yadav, V., Weiss, R. F., Duren, R. M., and Miller, C. E.: Carbon dioxide and methane measurements from the Los Angeles Megacity Carbon Project – Part 1: calibration, urban enhancements, and uncertainty estimates, Atmos. Chem. Phys., 17, 8313-8341, 10.5194/acp-17-8313-2017, 2017.

Wang, Y., Li, J. Y., Jing, H., Zhang, Q., Jiang, J. K., and Biswas, P.: Laboratory Evaluation and Calibration of Three Low- Cost Particle Sensors for Particulate Matter Measurement, Aerosol Science and Technology, 49, 1063-1077, 10.1080/02786826.2015.1100710, 2015.

Whelpdale, D. M., Low, T. B., and Kolomeychuk, R. J.: Advection climatology for the east coast of North America, Atmospheric Environment (1967), 18, 1311-1327, 10.1016/0004-6981(84)90040-4, 1984.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z.: Ensemble Data Assimilation with the NCEP Global Forecast System, Monthly Weather Review, 136, 463-482, 10.1175/2007mwr2018.1, 2008.

World Meteorological Organization: GAW Report No. 213: 17th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2013), http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_213_web.pdf (last access: July 11, 2016), 2013.

Wu, K., Lauvaux, T., Davis, K. J., Deng, A., Coto, I. L., Gurney, K. R., and Patarasuk, R.: Joint inverse estimation of fossil fuel and biogenic CO_2 fluxes in an urban environment: An observing system simulation experiment to assess the impact of multiple uncertainties, Elem Sci Anth, 6, 2018.

Wu, L., Broquet, G., Ciais, P., Bellassen, V., Vogel, F., Chevallier, F., Xueref-Remy, I., and Wang, Y.: What would dense atmospheric observation networks bring to the quantification of city CO_2 emissions?, Atmos. Chem. Phys., 16, 7743-7771, doi: 10.5194/acp-16-7743-2016, 2016.

Yasuda, T., Yonemura, S., and Tani, A.: Comparison of the Characteristics of Small Commercial NDIR CO_2 Sensor Models and Development of a Portable CO2 Measurement Device, Sensors, 12, 3641-3655, doi: 10.3390/s120303641, 2012.

Young, D. T., Chapman, L., Muller, C. L., Cai, X. M., and Grimmond, C. S. B.: A Low-Cost Wireless Temperature Sensor: Evaluation for Use in Environmental Monitoring Applications, Journal of Atmospheric and Oceanic Technology, 31, 938-944, 10.1175/jtech-d-13-00217.1, 2014.

Zeng, N., Mariotti, A., and Wetzel, P.: Terrestrial mechanisms of interannual CO_2 variability, Global Biogeochemical Cycles, 19, 2005.